

A Collective Assessment on Data Manipulation in Research Science

Kiran

B.Tech(EE), PGDESD, M.Tech(EE), PGDET

Research Scholar

Uttarakhand, India

Abstract –

Research covers a multidisciplinary field, growing everyday that involves wide range covering engineering, medical, legal, social and political fields which are unique and complex. A medical trials of patients could affect polices a government issues; A report on electricity usage could lead to companies taking step to manage its tariff charges. An honest research could help in many ways but there is a grey area of illicit, fraudulent practices. However, the question is where the lines must be drawn to avoid scientific misconduct and how it be judged? Researchers are central to this whole issue. Sometimes due to lure, to fall along the competing world some researcher give up to the burden hence data's are intentionally manipulated, cropped, morphed and values changed. This paper reviews the various data manipulative methods in research that could be used, it draws a collective assessment on alteration of data results or their interpretation in form of fabrication and falsification, intentional non publication of results, biased methodology, misleading reporting, plagiarism, professional misconduct. Concentration is made towards the Engineering Field.

Keywords: *Self citation, Plagiarism, Salami Slicing, Manipulation, Engineering, Electrical, p-hacking.*

INTRODUCTION

Technological breakthrough especially over the last 20 years, such as social networks, are making possible to pervasively capture huge amounts of data from many different contexts. This availability has made sharing and analyzing of research outcomes much easier. Increasingly, research studies are being conducted using existing data sets. Many research questions can be answered quickly and efficiently using data or specimens that have already been collected [9]. It is observed that the importance of intellectual property right have impinged on the academic community but the real potential lies in inhibiting the free flow of information that benefits new ideas, reproducibility, research data expansion, validation and transparency which is essential for scientific research to carry forward and flourish [52]. In particular, using an existing research outcome/data set can help a researcher obtain results more quickly, at a lower cost. Some benefits of shared 'research data' can have on the scientific communities are: First, Use of an existing data set tends to save a great deal of time, as opposed to designing a new research protocol to recruit subjects and gather data. This benefit of obtaining results more quickly is especially true for longitudinal studies like [50] used the data sets of [17] and [64] in which subjects

need not only be recruited, but also followed over months or years. Another example of benefits of data sharing is use of [30] as an inspiration to carry the research forward by [4], [62], [61]. Second, it tends to be much less expensive to use data that has already been collected [59]. Third, new research hypothesis can be tested more quickly [10] and data can be aggregated to study otherwise intractable issues. Fifth, other researchers can analyze data to explore the options for reproducibility beyond those who collected the original data [43] [68]. Recently there has been a welcome move to realign software engineering with different research fields as an evidence-based practice [6]. Many research groups are actively conducting empirical research [19] e.g. to compare different fault prediction models, Diagnosis and Maintenance planning of Electrical Substations [58] or evidence based environmental studies based on noise effects on high voltage transmission development [56]. Emerging Techniques like the systematic review [13] and meta-analysis [37] provide tools for standardized and sound assessment of the evidence available on different topics it can be viewed as the transference of good analytic practice from the single-study to the multiple- study context , some ranging examples are [51] [44] [28] [41]. However, this brings some challenges. First, for a particular question, how can we locate all the relevant evidence (primary studies) and make sense of them in an unbiased way. Second, what if some of these primary studies are inconsistent? [36] Traditionally believed to be rare or nonexistent. In which case how do we determine the 'true' answer? [69]. There is a difference between Honest statistical/Computing Error, Honest misunderstanding of data, Honest Misapplication of methods and then there is Malicious manipulation. Bad data practice is leading to bad research. The reliability of research is coming under increasing scrutiny [2]. This paper studies the motivating factors leading to use of data manipulation. It draws a collective outlook on major types of data manipulation practices being carried in research fields with main emphasis on engineering research.

MOTIVATION FOR MANIPULATION PRACTICES

Publication rates are increasing in line with accelerating scientific progress that is boosted by buoyant funding and advances in facilitating technologies. It is estimated 33100 active scholarly peer- reviewed English language in mid-2018 (plus a further 9400 non English Language journals) [54], collectively publishing over 3 million articles a year. Researchers write to keep records of their work for themselves, but more importantly for readers and peers, researchers willing to carry the work forward. [70] has

discussed the pragmatic reasons for writing up and publishing research results ranging in wanting to progress scientific thought; wanting to reach broad audience; improving chance of promotion.

Frequent publication is one of the few powerful methods at scholar's disposal to demonstrate academic talent to peers. Successful publication of research brings attention to scholars and their institutions; this in turn may bring in more funding for the institute and also ensure an individual's progress through their field. Academic institutions and university frequently use the number of publication to an individual's credit as the measure of competency [48]. Administrators are increasingly using this as the criteria during recruitments. Scholars, who publish infrequently or who focus on activities that does not result in publications like instructing undergraduates, may find themselves out of contentions for many teaching positions. It is due to these reasons that there is an immense pressure to publish. The phrase "Publish or perish" initially coined by Coolidge [54] in 1932 is now becoming a harsh reality.

The emphasis on publishing has decreased the value of the resulting scholarship as scholar must spend time scrambling to publish whatever they can manage, rather than spend time developing significant research agenda. The pressure to publish-or-perish also detracts from the time and effort professors can devote to teaching undergraduate and post-graduates. The rewards for exceptional teaching rarely match the rewards for exceptional research [55], which encourages faculty to favor the latter whenever they conflict. Many universities do not focus on teaching ability when they hire new faculty and simply look at the publications list [25]. This single-minded focus on the professor-as-researcher may cause faculty to neglect or be unable to perform some other responsibilities. This pressure to increase the number of publications has led to unethical practices and waste full research.

LOOPHOLES TO MANIPULATION PRACTICES

A journal paper reports on a finished piece of research or some significant achievement or a discovery made in a certain field of science [38]. [8] Questioned about how many duplicate and fake publications are in scientific Literature. In 2009, a systematic survey was carried by [12] in which scientists admitted to have fabricated, alteration in findings, falsified results, modified data- a serious misconduct that puts stake on the image of scientists as objective seekers of truth the display of lack of integrity is damaging the reputation of the fraternity as whole. A software method was provided by [8] to detect duplicate and fake publications. However, there is no universal procedure for checking the technical accuracy because the research techniques that researchers employ are varied as the fields themselves. Verification strategies to ensure both reliability and validity, some broadly accepted practices in fields of engineering, experimental observations must be shown to be reproducible in order to be creditable, checking, rechecking data to ensure that the interpretation is valid and also submitting the results to peer review. Peer review is regarded as the gold standard for research evaluation. Peer review is the stepping stone to publication, Peer review helps

validate research, establish a method by which it can be evaluated and increase networking possibilities within research communities. Many times peer review process is being compromised which has led to increased incidents of retracted papers [40]. A report [31] emphasizes on research reproducibility. Usually the published articles omits the parameters used, values, function, invocation sequences and other computational details which are critical for replicating results. [31] [27] [16] shared the necessity of data sharing and called for data and code be made available so that published computational results could be replicated. Nevertheless, the time-lag from bench to journal is decreasing and the pressure to publish mount with the increased chance of duplicated research and competitiveness. The following section represents a collection of data manipulation practices and its forms.

The Extent and Consequences of Data Dredging

Large amount of data is available in information industry and it is exceeding day by day. Data has evolved from a pure byproduct of value creation processes to a strategic resource and – in some cases – to a product in its own right. The growth of data and database has resulted in development of new technologies and tools to process the data into useful information [66]. The means and approaches to manage data are currently undergoing significant changes. Data mining covers vast research fields [39] [45] [23] [67]. Data Mining requires Data preprocessing which includes stages of data cleaning, data integration and data reduction [42] at this stage one of the types of bias which could be done is data dredging also called data snooping /data fishing it refers to the practice of misusing data mining techniques to show misleading scientific research [15]. Data snooping occurs when a given set of data is used multiple times for the purpose of inference or model selection this increases the possibility that any satisfactory result obtained is due to chance rather than any merit inherent in the method yielding the desired results. There the point to be noted is Data Snooping could be done professionally and ethically, or misleadingly and unethically, or misleadingly out of ignorance. Data snooping misleadingly out of ignorance is a common error in using statistics. The problems with data snooping are essentially the problems of multiple inferences [22]. Another common term for data dredging in scientific research is P-Hacking [5] described as a form of data manipulation wherein only selected data is included so to obtain a statistically significant result usually instigated with an aim to obtain positive result prioritizing over the actual observations. The urge for novel, confirmatory, and statistically significant results leads to substantial bias in the scientific literature.

Misappropriation of Credit

A published paper is a pathway to communicate research with the world. The archetypal image of a researcher publishing study findings as a lone author is long passé. Research is now primarily a collaborative and often an interdisciplinary endeavor, this shift has echoed a pattern of scientific

publications in which the author list has lengthened significantly often termed as 'mass authorship' or 'kilo-authorship' with some papers having thousands of authors. [24] sees a motivation factor in rise of shared authorship because Research Authorship not only confers professional advancement in academia it is also a yardstick for scholarship and promotion. However an author [18] is a person who provides a substantial intellectual contribution in research planning, investigation, data acquisition analysis and writing. Presumably, if a researcher is a part of collaborative team and does not participate in team's writing process, reviewing and approving the final version of the manuscript a situation may be compromised where intellectual contribution and accountability of author contribution becomes questionable [3]. In addition, one could drive a message that huge manpower and effort is needed to replicate the research and carry it forward. [29] Warned on the rise in 'guest-authors' and 'gift authorship' which is causing the academia to rethink how it assesses the worth of academic publications. References play a crucial role in scholarly and scientific writing for they allow the reader to explore in more detail a given line of thinking or evidence. For these reasons, it is important that authors strive for accuracy when listing references in manuscripts [11]. Unfortunately, it appears that some authors intentionally cite their own work regardless of its relevance, in an attempt to manipulate their own articles impact factor the consequence of such misuse can be profound [33]. Ghost authors are professional writers who provide substantial inputs but do not share authorship [47]. They are usually "pay-for-hire" writers using a writer who was not involved to generate the manuscript becomes murkier as it raises a doubt about the extent to which the external person had made substantive changes to the research work for the sake of grammatical assistance [24].

Lack/ Intended Omission of Negative Results

The Onus of research publication is to demonstrate originality. There is increasing concern that most published findings are on Positive Findings, the audit culture emphasis on positive findings by systematically excluding the negative finding/null results [53]. Many experimental results are never published for a large number of these come down to the data being negative and publishing negative findings are discouraged as they don't attract attention or citations and so they threaten journal impact factors [46]. But is this really the best approach to scientific results? What is the level of acceptability of negative results in research? Publishing positive results skews the scientific literature by only including chosen pieces of information [1] termed negative results as important, if not reported a non productive or flawed concept may continue causing a huge waste of time and resources, as other researcher considering the same questions may perform the same experiments. [49] pointed out that sharing negative findings can be valuable to the researchers in a number of ways, for example, by helping them to avoid repeating the same experiments, readjusting their research plans as well as encouraging new hypothesis building.

Unethical use of Salami Slicing

One of the agreeable aspects of editing a research paper is dealing with publication ethics. In recent times, the usage of least publishable unit has gained pace termed as salami slicing. Salami Slicing consists of splitting data from the same research into small units, each of which is published separately. This practice is neither new nor entirely culpable [26]. The question: When is Salami Slicing justified? If a research work is so extensive that it involves several research segments across discipline then dividing the research into several papers should not be criticized, as some topics could be too large for a single publication [21]. For example, studies that are performed using longitudinal data with several outcomes may warrant independent publication here it is the responsibility of author to ensure there is no overlap between research papers; in this case it is prudent to declare the journal editor regarding the previous publication at the time of submission that the data set presented in that paper is a part of larger study but when a researcher borne to the trap of 'publish or perish' employs salami slicing to increase their volume of publications it becomes an unethical practice. Salami Slicing may allow a researcher to progress of greater number of publication they can enlist but in long term the cost of unethical use of salami slicing diminishes the value of each publication, it also encourages fabrication of data, increases redundancy and unnecessary extrapolations of results [35].

Research Misconduct

Inaccurate data in scientific papers could be a result from honest error or intentional misconduct. Data fabrication occurs when actual study is not carried instead a fake data is used [57]. Data falsification is seen where the researcher did the experiment, but then changed some of the data [2]. Often termed as Data beautification in which data is altered, omitted, manufactured or misrepresented in a way that fits the desired outcome. Plagiarism threatens the foundations of knowledge. [65] [42] [34] have addressed and discussed plagiarism as a manipulative research practice that affects research Integrity. In general term, when an original idea of a published research work is used as source of inspiration without proper citation or sharing the credit it is considered as plagiarism. The frequency of plagiarism [60] is increasing and development of information and communication technologies facilitates it nowadays rarely plagiarism is a copy paste of a research work it ranges widely like borrowing subsets from previous publications, composing collages of works from different genres, changing the perspectives of the same research. Simultaneously, thanks to the same technology, plagiarism detection software is developing [63] [20]. There are several possible approaches for finding similarities in a given base of documents [7]. Quality is assessed on the basis of adequate evidence, while best results of the research are accomplished. Plagiarism Induces Pointless Research. It obstructs the work of scientific journals fighting plagiarism leading to waste of resources and time [32] [14].

CONCLUSION

This review focuses on behaviors that can falsify or bias scientific knowledge. The point, the problem is not confined to a bias published result. In many academic researches, grants/ funding are involved to carry a research forward. Thus data manipulated in an individual study could lead to other researchers exploring the same hypothesis; this will result in a significant waste of time and money. Further intentionally manipulated data obtained from different individual studies could drastically change results of systematic reviews and meta- analysis. A research outcome driven from manipulated data may lead to faulty decisions. Lower research quality, falsification, manipulation could affect the research productivity lead to wastage of opportunity, manpower and effort. To help sustain the reliability of Research a comprehensive assessment is needed to cross check discrepancies, greater policy relevance in scientific journals should encourage a systematic, objective, transparent means of evaluating and synthesizing evidence to sustain researcher ethics.

REFERENCES

- [1] A. Borji, "Negative Results in Computer Vision: A Perspective", *Image and Vision Computing*, 69: 1–8, 2018.
- [2] A. L. Zietman and Fastro, "Falsification, fabrication, and Palgiarism: The unholy trinityof scientific writing" *International journal of Radiation Oncology, biology, Physics*, 87(2):225-227, 2013.
- [3] A. Marusic, L. Bosnjak and A. Jeronic, "A Systematic review of research on the meaning. Ethics and practices of authorship across scholarly disciplines", *PLoS ONE*, 6(9): 1-17, 2011.
- [4] A. Mesbahi, M. Khafallah, A. Saad and A. Nouaiti, "Emulator design for a small wind turbine driving a self excited induction generator", *Third International Conference on Electrical and Information Technologies (ICEIT)*, 2017.
- [5] A. Nartey, "P-Hacking: its implication for science and scientific research", *Advances in Ophthalmology & Visual System*, 8(2): 115-116, 2018.
- [6] B. Kitchenham, O. P. Pearl Brereton, D. Budgen, M. Turner, J. Bailey and S. Linkman, "Systematic literature reviews in software engineering—A systematic literature review." *Information and Software Technology*, 51(1), 7-15, 2009.
- [7] C. Basile, D. Benedetto, and E. Caglioti, "A plagiarism detection procedure in three steps: Selection, Matches and "Squares", *Proceedings SEPLN Donostia, Spain*, 19-23, 2009.
- [8] C. Labbe and D. Labbe, "Duplicate and fake publications in scientific literature: how many SCI gen papers in computer science?" *Scientometrics*, Springer Verlag, 2012.
- [9] C. Wiley, "Data Sharing and Engineering Faculty: An Analysis of Selected Publications", *Science & Technology Libraries*, 37:4, 409-419, 2018.
- [10] C.L. Borgman, "Research Data: Who will share what, with whom, when, and why?", China-North America Library Conference (Beijing) Working Paper no.161, German Data Forum(RatWD), 2010.
- [11] D. A. Pendlebury, " The use and misuse of journal metrics and other citations indicators", *Archivum Immunologiae et therapac Experiments*, 57(1): 1-11, 2009.
- [12] D. Fanelli, "How many scientists fabricate and falsify research? Asystematic review and meta- analysis of survey data", *PLoS ONE*, 2(5):1-11, 2009.
- [13] D. Gough, S.Oliver and J.Thomas, "An Introduction to Systematic reviews" *SAGE Publications*, Second Edition, ISBN: 9781473929432, 2017.
- [14] D. Isoc, "Preventing Plagiarism in Engineering Education Research" , *Conference Proceedings International Symposium on Fundamentals of Electrical Engineering, IEEE*, 1-7, 2014.
- [15] D. Jensen, "Data Snooping, Dredging and Fishing: The Dark Side of Data Mining A SIGKDD99 Panel Report", *ACM SIGKDD Explorations Newsletter*, 1(2): 52, 2000.
- [16] D. L. Donoho, A. Maleki, I. U. Rahman, M. Shahram and V. Stodden, "Reproducible Research in Computational Harmonic Analysis", *Computing in Science & Engineering*, 11(1), 8–18.
- [17] D. Murray, L. Stankovic, V. Stankovic, "An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study", *Sci Data* 4:160122, 2017.
- [18] E,E. Tarkang, M. Kweku and F. B. Zotor, "Publiction practicesand responsible authorship: A Review Article", *Journal of Public Health in Africa*, 8(23): 36-42, 2017.
- [19] E. Bertino, "Introduction to Data science and Engineering", *Data Science and Engineering*, 1(1):1, 2016.
- [20] E. M. Bucci, "Automatic detection of Image Manipulations in the biomedical Literature", *Official Journal of the Cell Death Differentiation Association (Springer)*, 9:400, 2018.
- [21] Editorial, "Duplicate publication and 'salami slicing': Ethical issues and practical solutions", *International Journal of Nursing Studies (Elseiver)*, 45: 1257-1260, 2008.
- [22] H. White, "A Reality check for Data Snooping", *Econometrica*, 68(5): 1097-1126. 2000.
- [23] I. Colak, S. Sagiroglu and M. Yesilbudak, "Data miningand Wind Power prediction: A literature

- Review”, *Renewable Energy*, 46: 241–247, 2012.
- [24] J. Borenstein and A. E. Shamoo, “Rethinking Authorship in the era of collaborative Research”, *Accountability in Research*, 22(5): 267-283, 2015.
- [25] J. D. Linton, R.Tierney and S.T. Walsh, “Publis or Perish: How are research and reputation related?”, *Serials Review*, 37:244-257, 2011.
- [26] J. Karlsson and P. Beaufiles, “Legitimate division of large data sets, salami slicing and dual publication, where does a fraudbegan?”, *Springer*, 21(4):751- 752, 2013.
- [27] J. Kovacevic, “How to Encourage and Publish Reproducible Research”, *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, 2007.
- [28] J. Minton and E. Brambley, “Meta- Analysis of curvature trends in asymmetric rolling”, *International Conference on the Technology of Plasticity*, ICPT, Procedia Engineering 207: 1355-1360, 2017.
- [29] J. Schofferman, “Ghost and Guest Authors: You can’t always trust who you read”, *Wiley Periodicals*, 16: 416-420, 2015.
- [30] J.Vaheeshan, V.Vihirithanath, S. G. Abeyaratne, A.Atputharajah and G. Ramatharan, “Wind Turbine Emulator”, *Sixth International Conference on Industrial and Information Systems*, pp. 511-516, 2011.
- [31] John Bailleul, “Report on the First IEEE Workshop on th future of Research Curation and Research Reproducibility”, *IEEE*, Washington, USA, 2016.
- [32] K. Bakhtiyari, H. Salehi, M. A. Embi, M. Shakiba, A. Zavvari, M. S. Moghadam, N. A. Ebrahim and M. Mohammadjafari, “ Ethical and Unethical Methods of Plagiarism Prevention in Academic Writing”, *International Education Studies*, 7(7): 52-62, 2014.
- [33] K. Hyland, “Self-citation and self- reference: credibility and promotion in academic publication”, *Journal of the American Scoiety for Infrmation Science and Technology*, 54(3): 251-259, 2003.
- [34] L. K. John, G. Loewenstein and D. Prelec, “Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling”, *Psychological Science*, 23(5): 524-532, 2012.
- [35] L. Souder, “The ethic of scholarly peer review: a review of the literature”, *learned Publishing*, 24(1):55-74, 2011.
- [36] L.Consoli, “Scientific misconduct and science ethics: a case study base approach”, *Science and Engineering Ethics*, 12(3): 533-541, 2006.
- [37] M. Borenstein, L. V. Hedges, J.P.T. Higgins and H.R. Rothstein, “Introduction to Meta- Analysis”, *John Wiley & Sons*, ISBN: 9780470743386, 2009.
- [38] M. Derntl, “Basics of research paper writing and publishing”, *International Journal Technology Enhanced Learning*, 6(2):105-123, 2014.
- [39] M. Kazerooni, H. Zhu and T. J. Overbye, “Literature Review on the applications of Data Mining in Power System”, *Power and Energy Conference at Illinois (PECI)*, 2014.
- [40] M. L. Grieneisen and M. Zhang. “A comprehensive survey of retracted articles from the scholarly literature” , *PloS ONE*, 7(10):1-15, 2012.
- [41] M. Mattmann, I. Logar and R. Brouwer, “Wind Power Externalities: A meta-analysis”, *Ecological Economics*, 127: 23-36, 2016.
- [42] M. Roig, “Avoiding plagiarism, self- plagiarism, and other questionable writing practices: A guideto ethical writing”, *New York: St. Johns University Press*, 2006.
- [43] M. Schwab, N. Karrenbach, and J. Claerbout, “Making scientific computations reproducible”, *Computing in Science & Engineering*, 2(6): 61–67, 2000.
- [44] M. Shepperd, “Combining Evidence and Meta-analysis in Software Engineering”, *Lecture Notes in Computer Science*, vol. 7171, Springer, Berlin, Heidelberg, ISSSE 2010, ISSSE 2009, ISSSE 2011.
- [45] M. Yesilbudak, M. Colak and R. Bayindir, “A Review of Data Mining and Solar Power Prediction”, *5th International Conference on renewable Energy Research and Applications*, 2016.
- [46] N. Matosin, E. Frank, M. Engel, J. S. Lum and K. A. Newell, “Negativity towards results: A discussion of the disconnect between scientific worth and scientific culture”, *The company of Biologists Ltd*, 7: 171-173, 2014.
- [47] P. C. Gøtzsche, A. Hróbjartsson, H. K. Johansen, M. T. Haahr, D. G. Altman and An-Wen Chan, “Ghost Authorship in Industry- Initiated Randomised Trials”, *PLoS Medicine*, 4(1):47-52, 2007.
- [48] P. F. Ribeiro, “Publish or perish: An evaluation of the quality, quantity, ethics and review process of IEEE/PES publications: A summary of the Power Globe discussion”, *IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, 2008.
- [49] P. G. Weintraub, “The Importance of Publishing Negative Results”, *Journal of Insect Science*, 16 (1): 109, 2016.
- [50] P. Huber, M. Gerber, A. Rumsch and A. Palice, “Prediction of domestic appliances usage based on electrical consumption”, *energy Informatics*, 1(10):265, 2018.
- [51] R. Bernards, J.Reinders, E. Klaassen, J. Moreen and H. Slootweg, “Meta- Analysis of the results of

- European smart-grid projects to quantify residential flexibility”, *CIREC Workshop- Helsinki*, Paper-0342, 2016.
- [52] R. Elliott, “Who owns scientific data? The impact of intellectual property rights on the scientific publication chain” *Learned Publishing*, 18:91–94, 2005.
- [53] R. F. Paige, J. Cabot and N. A. Ernst, “Foreword to the special section on negative results in Software Engineering”, *Empirical Software Engineering*, 22(5):2453-2456, 2017.
- [54] R. Johnson, A. Watkinson and M. Mabe, “*STM Report- An overview of scientific and scholarly publishing*” fifth edition, 2018.
- [55] R.Plomp, “The Highly cited papers of Professors as an indicator of a research groups scientific performance”, *Scientometrics*, 29(3): 377-393, 1994.
- [56] RPS Group, “EriGrid Evidence based Environmental Studies Study 8: Noise”, 2016
- [57] S. Bredl, N. Storfinger and N. Menold, “ A Literature Review of Methods to Detect Fabricated Survey Data”, *Discussion Papers from Justus Liebig University Giessen, Center for international Development and Environmental Research (ZEU)*, No 56, 2019.
- [58] S. Costinas, Gh. Comanescu, I. Fagarason, “ Monitoring, Diagnosis and Maintenance planning for equipments in electrical substations” *The Sixth World Energy System Conference (Torino, Italy)*, Vol: Proceedings of the sixth world energy system, Paper WA 151, 405-410, 2006.
- [59] S. D. Tiessen, R. Darby, K. Gitmans, S. Lambert, B. Matthews, S. Mele, J. Suhonen and M. Wilson, “Enabling Sharing and Reuse of Scientific Data”, *New Review of Information Networking*, 19(1), 16–43, 2014.
- [60] S. K. Sterba, “Misconduct in the Analysis and Reporting of Data: Bridging Methodological and Ethical Agendas for Change”, *Ethics & Behaviour*, 16(4), 305-318.
- [61] S. Kouadria, S. Belfedhal, E.M. Berkouk and Y. Meslem, “Development of real time wind turbine emulator based on DC motor controlled by PI regulator”, *Eighth International Conference and Exhibition on Ecological Vehicles and Renewable Energies (EVER)*, 2013.
- [62] S. Kouadria, S. Belfedhal, Y. Meslem and E.M. Berkouk, “ Development of real time Wind Turbine Emulator based on DC Motor controlled by hysteresis regulator”, *International Renewable and Sustainable Energy Conference (IRSEC)*, 2013.
- [63] S. M. Alzahrani, N. Salim and A. Abraham, “Understanding Plagiarism LinguisticPatterns, Textual Features, and Detection Methods”, *IEEE Transactions on Systems, Man and Cybernetics- Part C Applications and Reviews*, 42(2): 133-149, 2012.
- [64] S. Makonin, B. Ellert, I.V. Bajić and F. Popowich, “Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014”, *Scientific Data*, 3, 160037, 2016.
- [65] S. Pandey, H.R. Sharma and A. Rawal, “A Review Paper on Awareness Statistics on Palgiarism among Research Scholars”, *IEEE International Conference on Recent Trends in Electronics Information Communication Technology*, 24-26, 2016.
- [66] S.J.Lee and K.Siau, “A Review of data mining techniques”, *Industrial Management & Data Systems*, 101(1):41-46, 2001.
- [67] T. O. Aro, B. J. Muhammed, O. B. Ayoade and I. D. Oladipo, “A Review on data Mining Techniques for Heart Disease redictions”, *Annals, Computer Science Series*, 15th Tome 1st fase, 99-103, 2017.
- [68] V. Stodden, “The Scientific Method in Practice: Reproducibility in the Computational Sciences.” *SSRN Electronic Journal*, 10.2139/ssrn.1550193, 2010.
- [69] V.Pupovac and D. Fanelli, “Scientists admitting to Palagarism: A Meta- Analysis”, *Science and Engineering Ethics*, 21(5):1331-1352, 2015.
- [70] Z. Šesták, J. Peat, E. Elliott, L. Baur and V. Keena “Scientific Writing. Easy When You Know How” *Photosynthetica*, 41(3):392, 2003.