

Review on INArxiv preprint: K-Means Method for Clustering Water Quality Status on The Rivers of Banjarmasin

Dasapta Erwin Irawan¹ and PREreview Team²

¹Institut Teknologi Bandung

²Affiliation not available

February 17, 2018

Abstract

This is a preprint journal club review of K-Means Method for Clustering Water Quality Status on The Rivers of Banjarmasin by Tien Zubaidah and Nieke Karnaningroem. The preprint was originally posted on INArxiv on December 21, 2017 (link: <https://osf.io/g6wkp/>). The article is now in review in the ARPJ Journal of Engineering and Applied Sciences (submitted December 20, 2017).

Original abstract: The surface river water quality in Banjarmasin city tends to decline constantly as the result of direct and indirect waste disposal from various human activities along the river body. This study aimed to determine the vulnerability points against pollution in the rivers of Banjarmasin using clustering techniques with K-means algorithm. The parameters observed include Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Total Suspended Solid (TSS) and Dissolved Oxygen (DO). The data were collected at eight water monitoring stations on various rivers in Banjarmasin city. With the K-means method, four water quality status were clustered. The result showed that 6 stations observed during the period April to October 2016 were categorized into the heavy polluted cluster with major pollution point of sources came from the domestic and industrial activities.

Short review

Introduction

Yang terhormat Tim Penulis (Ibu Tien dan Ibu Nieke)

Pertama-tama saya mengucapkan terima kasih telah menggunakan layanan INArxiv. Semoga layanan kami ini dapat memaksimumkan pengembangan risetnya. Ulasan ini juga menjadi bukti bahwa salah satu manfaat mengunggah naskah preprint adalah mendapatkan umpan balik dari rekan sebidang. Saat ini cara untuk memberikan ulasan pun tidak konvensional seperti dulu, melalui editor sebuah jurnal, melainkan bisa melalui saluran seperti blog pribadi, e-mail, atau bahkan ulasan publik melalui laman *journal club* seperti ini.

Reviewing has been extensively developed from conventional way, via journal editor, to a more open and personal way directly to the author, via personal email, personal blogpost, or via a journal club like this review.

Saya tertarik dengan naskah berikut ini. Sejalan dengan fokus riset saya di bidang hidrogeologi. Makalah ini dengan sangat baik telah menjelaskan alur riset dengan sangat baik. Hasilnyapun telah merepresentasikan

tujuan riset. Penurunan kualitas air sungai sudah menjadi masalah di mana-mana, terutama di kota besar di dunia. Berbagai metode memang sudah seharusnya digunakan untuk memerangi hal tersebut. Salah satunya adalah metode karakterisasi kualitas air dengan teknik *K-Means clustering*.

| River contamination has been a major problem in big cities in Indonesia and also the world

Pentingnya *reproducibility* dalam sains

Namun demikian perkenankan saya menambahkan sedikit catatan untuk makalahnya. Di dunia saat ini sedang berkembang konsep sains terbuka (*open science*) yang salah satunya mendorong setiap makalah perlu memiliki struktur untuk dapat direproduksi. Reproduksi (produksi ulang) sangat penting dalam dunia saintifik, karena setiap output riset harus dapat dianalisis ulang oleh pihak lain untuk menjamin validitasnya. Makalah yang *reproducible* juga memiliki peluang disitat lebih besar. Dua komponen penting untuk mendukung ini adalah: data mentah dan metode analisis yang rinci.

Yang saya pahami, sering terjadi penulis terkendala oleh jumlah halaman maksimum atau jumlah kata maksimum untuk dapat lebih rinci menjelaskan dua hal di atas. Untuk data mentah akan bagus kalau diunggah. Dalam kasus makalah ini sedang direview oleh sebuah jurnal, maka data mentah dapat diunggah sekaligus dalam halaman preprint ini. Ibu dapat melakukan “edit preprint” dan mengunggah file data mentah dalam berbagai format (xls, csv, dll). Kemudian dalam naskah disunting dengan menambahkan tautan preprint di atas untuk mengarahkan pembaca yang berminat untuk mengetahui data. Berikut ini saya sampaikan diagram spektrum dalam *reproducibility* ([the](#)).

| Reproducibility is very important to ensure the validity of a research (or paper)

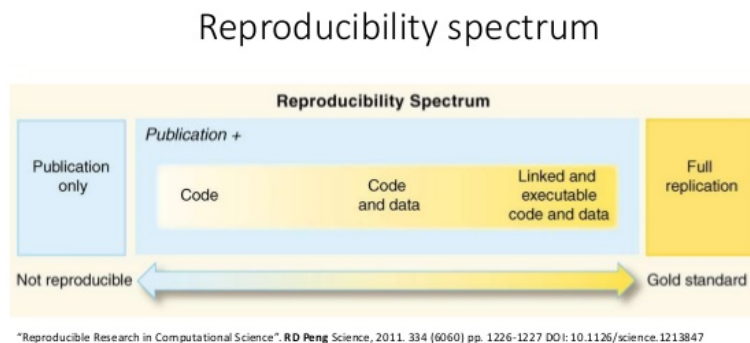


Figure 1: Spektrum *reproducibility* (Peng, 2011)

Bila ada pembaca yang berminat mereproduksi makalah atau akan menggunakan data mentahnya, maka ia dapat melakukan “forking” (pencabangan proyek) ini dengan mudah untuk masuk ke dalam proyek OSF

ybs. Kemudian ybs dapat melakukan berbagai analisis sesuai keperluannya tanpa mengubah file-file yang ada dalam proyek ibu.

| Someone who wishes to reproduce an OSF-based preprint could easily “fork” the original document to do his/her own analysis

Untuk analisis, Ibu dapat menjelaskan tahapan rinci pengolahan data. Akan lebih bagus kalau piranti lunak yang digunakan berbasis teks (*command line*), dengan demikian, ibu cukup melampirkan file kode sebagai file pendukung makalah secara terpisah (lihat gambar di atas).

Peluang pengembangan dalam analisis

Dalam analisis disajikan beberapa penyebab mengapa terjadi perubahan kluster pada bulan yang berbeda, namun penjelasan baru memasukkan variabel curah hujan. Akan lebih baik bila makalah ini dikembangkan dengan mengaitkan perubahan tersebut dengan perubahan tata guna lahan atau kontribusi output air tanah yang mengalir menuju sungai. Alasan saya mengemukakan ini adalah BOD atau COD biasanya berkaitan dengan limbah domestik atau peternakan yang membuang limbah langsung ke sungai tanpa pengolahan terlebih dahulu, yang mana ini sering terjadi di Indonesia. Perubahan tata guna lahan dari lahan terbuka menjadi perumahan bisa jadi salah satu indikasi.

Alasan saya yang kedua adalah, pada Sungai Cikapundung ([Darul et al., 2015](#)), Sungai Ciliwung ([Irawan et al., 2014](#)), dan Cisadane ([Dasapta Erwin Irawan, 2016](#)) serta anak-anak sungai di Gunung Ciremai ([Irawan et al., 2009](#)), sangat mungkin ada kontribusi air tanah ke sungai. Rembesan air dari akuifer ini juga dapat membawa zat polutan terutama bila aliran air tanah bersinggungan dengan titik-titik sumber kontaminasi organik, misal *septic tank*.

Pengamatan secara time series pada beberapa tahun dengan melibatkan analisis citra satelit ([Achmad Darul, 2017](#); [Rahmadia, 2017](#)) akan sangat bagus. Kekuatan analisis spasial dapat melengkapi analisis statistik yang telah dilakukan oleh penulis.

Peluang kolaborasi

Yang ketiga, saya berminat untuk melakukan kolaborasi menulis paper bersama Ibu dan rekan-rekan yang lain yang mungkin berminat. Usulan saya, kita gabungkan data mentah dari berbagai lokasi dan dianalisis secara bersama untuk mengetahui kesamaan atau perbedaan karakter diantaranya. Secara spasial saya juga ingin membuat clustering dengan metode hirarkis yang lebih visual. Bila itu digabungkan dengan analisis spasial, maka akan lebih bermanfaat bagi pihak pengelola daerah. Strategi pengelolaan air bisa sangat bergantung kepada hasil analisis tersebut.

| With this review, I hereby also propose to make a potential collaboration

Demikian kurang lebih komentar dan juga proposal saya agar dapat dimanfaatkan bila ada kesempatan.

Salam, Erwin (osf.io/he3j7)

[Institut Teknologi Bandung](#)

Personal blog

References

- Reproducible research: theory. <https://www.slideshare.net/CTobinMagle/reproducible-research-theory>. URL <https://www.slideshare.net/CTobinMagle/reproducible-research-theory>. Accessed on Sat, December 23, 2017.
- Raden Ajeng Koesoemo Roekmi Putu Oktavia Achmad Darul, Dasapta Erwin Irawan. PRELIMINARY STUDY NORMALIZED DIFFERENCE WATER INDEX IN THE RECHARGE AREA OF GROUND WATER OF BEKASI DISTRICT — Darul — *Jurnal Geografi Gea. Jurnal Geografi GEA Universitas Pendidikan Indonesia*, 2017. URL <http://ejournal.upi.edu/index.php/gea/article/view/8134>. Accessed on Sat, December 23, 2017.
- A. Darul, D. E. Irawan, and N. J. Trilaksono. Groundwater and river water interaction on Cikapundung River: Revisited. AIP Publishing LLC, 2015. doi: 10.1063/1.4930778. URL <https://doi.org/10.1063/2F1.4930778>.
- Defitri Yeni Arno Adi Kuntoro Miga Magenika Julian Dasapta Erwin Irawan, Deny Juanda Puradimaja. [1603.09380] Decreasing groundwater quality at Cisadane riverbanks: groundwater-surface water approach. *Arxiv*, 2016. URL <https://arxiv.org/abs/1603.09380>. Accessed on Sat, December 23, 2017.
- D. E. Irawan, H. Silaen, P. Sumintadireja, R. F. Lubis, B. Brahmantyo, and D. J. Puradimaja. Groundwater-surface water interactions of Ciliwung River streams segment Bogor-Jakarta, Indonesia. *Environmental Earth Sciences*, 73(3):1295–1302, jul 2014. doi: 10.1007/s12665-014-3482-4. URL <https://doi.org/10.1007/2Fs12665-014-3482-4>.
- Dasapta Erwin Irawan, Deny Juanda Puradimaja, Sudarto Notosiswoyo, and Prihadi Soemintadiredja. Hydrogeochemistry of volcanic hydrogeology based on cluster analysis of Mount Ciremai West Java, Indonesia. *Journal of Hydrology*, 376(1-2):221–234, sep 2009. doi: 10.1016/j.jhydrol.2009.07.033. URL <https://doi.org/10.1016/2Fj.jhydrol.2009.07.033>.
- R. D. Peng. Reproducible Research in Computational Science. *Science*, 334(6060):1226–1227, dec 2011. doi: 10.1126/science.1213847. URL <https://doi.org/10.1126/2Fscience.1213847>.
- Endah; Irawan Dasapta Erwin; Rosada Keukeu K. Rahmadia, Tantra; Sulistyawati. Cikapundung watershed land cover characterization using SPOT-6 imagery. In *Proceeding of ICMSE 2017*, 2017. URL https://figshare.com/articles/Cikapundung_watershed_land_cover_characterization_using_SPOT-6_imagery/5425729. Accessed on Sat, December 23, 2017.