Migration Letters

Volume: 21, No: S3 (2024), pp. 134-138

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)

www.migrationletters.com

Visual Data Analysis Applied To Biotechnology: A Case Study In Volatile Fatty Acids

Jeniffer Gracia[1], Carlos Montenegro [2], Jaime Páez [3], Paola Acevedo[4], Iván Cabeza[5]

Abstract

Visual data analysis allows information to be communicated in a graphical, interactive and understandable way and is an area of research with wide application in various fields. The visualization of multidimensional data sets, correlations, patterns and trends are taking on an increasingly important role in research. With traditional techniques, partial views of the data are generated; therefore, making an effective visualization becomes an increasingly complex task. In this context, the general objective of the study is to use new technologies in the field of biotechnology by applying a visual data analysis tool to the case study on the production of Volatile Fatty Acids (VFA) from the sludge of a Wastewater Treatment Plant (WWTP). VFA production was analyzed as a function of time, organic load, sludge type, pH, and temperature.

Keywords *Data Analysis, Visual Analysis, Data Visualization, Data Processing, Dashboard.*

Resumen

El análisis visual de datos, permite comunicar información de una manera gráfica, interactiva y comprensible y es un área de investigación con una gran aplicación en diversos campos. La visualización de conjuntos de datos multidimensionales, correlaciones, patrones y tendencias, toman un rol cada vez más importante en la investigación. Con las técnicas tradicionales, se generan vistas parciales de los datos, por lo tanto, la realización de una visualización eficaz se vuelve cada vez más compleja. En este contexto, el objetivo general de este trabajo, es utilizar estas nuevas tecnologías en el campo de[1] la biotecnología, aplicando una herramienta de análisis visual de datos para el caso de estudio de producción de ácidos grasos volátiles-AGV's, a partir de los lodos de una planta de tratamiento de aguas residuales. Se analizó la producción de AGV´s en función de las variables tiempo, carga orgánica, tipo de lodo, pH y temperatura.

Palabras clave

Análisis de Datos, Análisis Visual, Visualización de Datos, Procesamiento de Datos, Dashboard.

Introduction

Visual data analysis aims to foster effective collaboration between humans and machines to improve the knowledge generation process (Sacha et al., 2016). The large amount of data resulting from a research project has become one of the main challenges for those who must interpret results for decision-making, partly because they do not have the appropriate tools to

[1]Universidad Distrital Francisco José de Caldas, Colombia
[2] Universidad Distrital Francisco José de Caldas, Colombia.
[3] Instituto Técnico Central, Colombia.
[4] Universidad Cooperativa de Colombia
[5] Universidad de la Sabana.

analyze and interpret records, variables and indicators, and although data protection is required, it is essential to process and interpret them.

Among the current technological solutions that allow the treatment or analysis of data is data mining, which is an integral part of the discovery of knowledge, since it is a process that allows converting raw data into useful information (TAN et al., 2006).

Typically, data managers use widespread tools for the storage of their databases, such as office packages, such as Excel (.xlsx, .xls), files in .csv format (comma-separated values), files in .xml language (extensible markup language), files in .sql format (structured query language); however, the previous data formats are not entirely helpful to perform a visual analysis that allows the straightforward interpretation of the data.

The visual analysis process begins with collecting the data to be analyzed, which is then processed and finally produces a visualization that allows interaction with the user (Angelini et al., 2018). Visualization can convey information through graphical representations, making complex data more accessible, understandable and usable, allowing users to quickly analyze and reason about the data (Liu et al., 2014).

According to El-Assady et al. (2014), three main components are identified to describe the visual predictive analytics pipeline: data preprocessing, feature selection and generation, training, model selection and validation (Lu et al., 2017). For example, a significant challenge in data preprocessing is ensuring data quality and allowing data disputes (Kandel et al., 2011).

Visual analytics is particularly concerned with coupling interactive visual representations with data mining techniques that can be widely applied to analytical reasoning to support the sense-making process for various applications (Zhu et al., 2018). One of the main problems in establishing a visual data analysis model arises from correctly establishing the variables and data necessary to design the model (Gel et al., 2015). For this reason, it is necessary to consult with experts in the particular area in order to achieve a higher quality and fit-for-purpose response and to determine the information needed to build the model.

Taking into account that visual data analysis is nowadays widely applied in different fields of the environmental area, as stated by Haowen et al. (2022), the purpose of this contribution is to analyze the results obtained in biotechnological research, in which the production of volatile fatty acids was performed using sludge from El Salitre Wastewater Treatment Plant (WWTP), by applying the Microsoft Power BI visual data analysis tool.

Methodology

In order to analyze the data on the production of volatile fatty acids from El Salitre WWTP, using a visual analysis technique, the steps described in Figure 1 are applied.

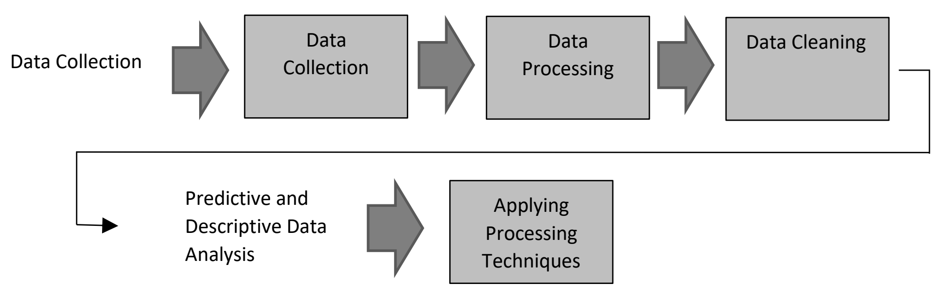

Figure 1. Research Methodology: Visual Data Analysis of Volatile Fatty Acids Production Data.

Results

The data obtained from the experiment results on the production of volatile fatty acids using sludge from El Salitre WWTP were compiled from a total of forty-eight experiments carried out at a laboratory scale in 200 ml reactors in triplicate. The bioproduction of volatile fatty acids was carried out with two types of sludge: digested sludge and primary sludge, which used organic loads of 4, 6, 10 and 14 g/SV (grams of volatile solids), respectively, using pH of 9.5, 10.5 and 11.5. Additionally, the experiments were conducted at different temperatures of 25°C, 35°C, 45°C, and 55°C, while the reactors were in production time from 2 to 12 days for a total of 576 VFA production reactors and 176 control reactors, that were not used in the present analysis. The total experimental data was 752 once the data was cleaned, with 232 for the primary sludge and 234 for the digested sludge.

Tests were made on the Tableau dashboard and Power BI. Finally, it was decided to perform the analysis on Microsoft's Power BI visual analysis tool since it is a tool with which the researchers have an official license, unlike Tableau software.

Figures 1 to 4 consolidate the analysis performed in the Power BI visual data tool for the combinations of digested sludge and primary sludge at the four temperatures used in the experiment.

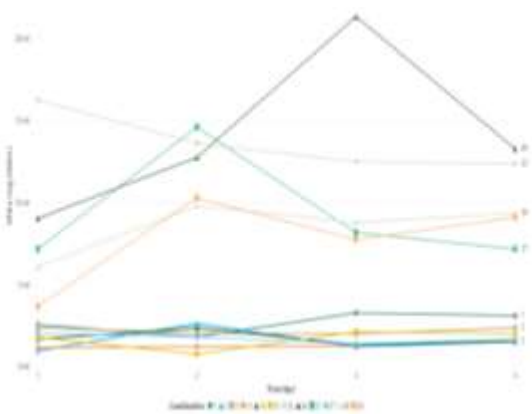

Figure 2. Digested sludge and primary sludge production at 25°C.

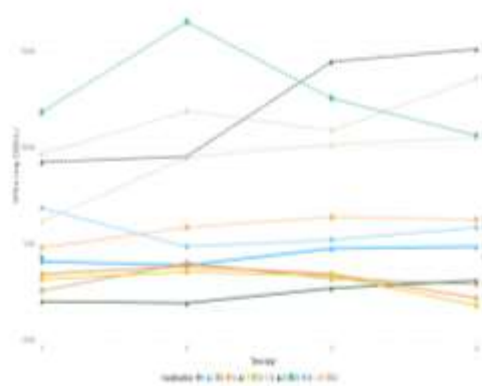

Figure 3. Production of digested sludge and primary sludge at 35°C.

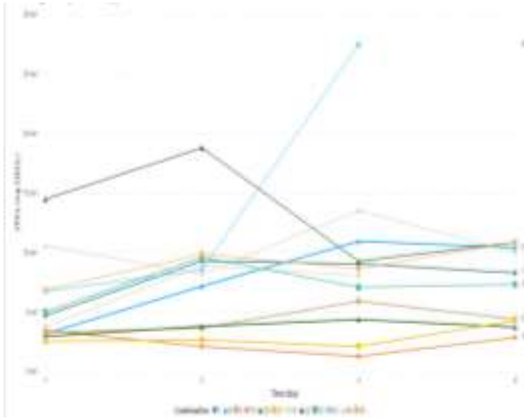

Figure 4. Production of digested sludge and primary sludge at 45°C.

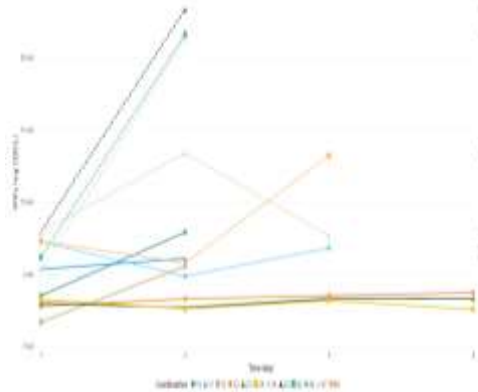

Figure 5. Production of digested sludge and primary sludge at 45°C.

The dotted lines correspond to the primary sludge, and the continuous lines correspond to the digested sludge; the bevels in a circle, triangle and square correspond to pH 9.5, 10.5 and 11.5, respectively. Likewise, the numbers assigned from 1 to 48 represent the number of the combination of the analyzed experiment. The Y-axis shows the VFA production in mg COD/L, and the X-axis is the retention time of the reactor.

Discussion and Conclusions

In this case study, data from investigating the bioproduction of volatile fatty acids using sludge from El Salitre WWTP as raw material was processed. The researchers obtained experimental results with data treatment and then analyzed them in the Microsoft Power BI tool.

Visual data analysis tools are very useful for organizing and studying research results since they provide various graphical options that support interactive visualization and allow observing what happens when variables are combined, leading researchers to a practical analysis of results and facilitating decision-making.

References

Angelini, M., Santucci, G., Schumann, H., & Schulz, H.-J. (2018). A Review and Characterization of Progressive Visual Analytics. Informatics, 5(3), 31. https://doi.org/10.3390/informatics5030031

Drumm, C., Schmitt, M., Do, H.-H., & Rahm, E. (2007). QuickMig -Automatic Schema Matching for Data Migration Projects Figure 1: Running example showing a mapping of customer data. 107–116.

El-Assady, M., Sevastjanova, R., Sperrle, F., Keim, D., & Collins, C. (2018). Progressive Learning of Topic Modeling Parameters: A Visual Analytics Framework. IEEE Transactions on Visualization and Computer Graphics, 24(1), 382–391. https://doi.org/10.1109/TVCG.2017.2745080

Gel, M., Urciuolo, A., & Iturraspe, R. (2015). Lenguaje Específico de Dominio para Aplicaciones de Negocios. Revista Eletrônica Argentina-Brasil de Tecnologias Da Informação e Da Comunicação, 1(3). https://doi.org/10.5281/zenodo.59457

Haowen, X, Andy, B, Yan, L, Melissa, R. Allen-D, Jibonananda, S. (2022). An overview of visualization and visual analytics applications in water resources management. Environmental Modelling & Software. 153. https://doi.org/10.1016/j.envsoft.2022.105396.

Kandel, S., Paepcke, A., Hellerstein, J., & Heer, J. (2011). Wrangler: interactive visual specification of data transformation scripts. ACM CHI Conference on Human Factors in Computing Systems, 3363. https://doi.org/10.1145/1978942.1979444

Krause, J., Perer, A., & Bertini, E. (2014). INFUSE: Interactive feature selection for predictive modeling of high dimensional data. IEEE Transactions on Visualization and Computer Graphics, 20(12), 1614–1623. https://doi.org/10.1109/TVCG.2014.2346482

Liu, S., Cui, W., Wu, Y., & Liu, M. (2014). A survey on information visualization: recent advances and challenges. Visual Computer, 30(12), 1373–1393. https://doi.org/10.1007/s00371-013-0892-3

Lu, J., Chen, W., Ma, Y., Ke, J., Li, Z., Zhang, F., & Maciejewski, R. (2017). Recent progress and trends in predictive visual analytics. Frontiers of Computer Science, 11(2), 192–207. https://doi.org/10.1007/s11704-016-6028-y

Sacha, D., Senaratne, H., Kwon, B. C., Ellis, G., & Keim, D. A. (2016). The Role of Uncertainty, Awareness, and Trust in Visual Analytics. IEEE Transactions on Visualization and Computer Graphics, 22(1), 240–249. https://doi.org/10.1109/TVCG.2015.2467591

TAN, P.-N., Steinbach, M., & Kumar, V. (2006). Pang-Ning Tan - Introduction to data mining. https://doi.org/10.1152/ajpgi.1999.276.5.G1279

Zhu, X., Wang, C., Gu, T., Liu, Y., Ren, Q., & Zhou, Z. (2018). VisUPI: visual analytics for University Personality Inventory data. Journal of Visualization, 21(5), 885–901. https://doi.org/10.1007/s12650-018-0499-x