

WhatsApp conversation analysis as a means of motivation for teaching in data science

Análisis de conversaciones de WhatsApp como medio de motivación para la enseñanza en ciencia de datos

NAVARRERE-ARIAS, Dulce J.†*, HERNÁNDEZ-GARCÍA, Héctor Daniel and PÉREZ-BAUTISTA, Mario

Instituto Tecnológico Superior del Occidente del Estado de Hidalgo, División de Ingeniería en Sistemas Computacionales, México

ID 1st Author: Dulce J., Navarrete-Arias / ORC ID: 0000-0002-7915-068X, Researcher ID Thomson: AAR-8785-2021, CVU CONACYT ID: 366071

ID 1st Co-author: Héctor Daniel, Hernández-García / ORC ID: 0000-0001-5261-8353, CVU CONACYT ID: 208146

ID 2nd Co-author: Mario, Pérez-Bautista / ORC ID: 0000-0002-3260-906X, CVU CONACYT ID: 638669

DOI: 10.35429/JMQM.2021.9.5.18.24

Received July 28, 2021; Accepted December 30, 2021

Abstract

Both social networks and instant messaging systems have grown in the last decade, becoming an important means of acquiring and disseminating information in different domains, such as business, entertainment, politics, among others. This case study aims to analyze the motivation for learning the discipline Data Science (DS) within the population of engineering students in computer systems in the seventh semester of the Technologic Superior del Occidente del Estado de Hidalgo (ITSOEH), the experiment it consisted of applying a survey in two moments, at the beginning and end of the Introduction to Data Science course within a Likert scale, this to measure the degree of motivation of the students in the area of SD. First, different WhatsApp conversations were obtained and then an analysis was carried out through the statistical language R and the RStudio development environment, the data was examined using graphs to visualize the behavior of the conversation participants. The results of the case study show that students are aware of the power of data for large companies, but are unaware of the use of data generated by our fingerprint when browsing the internet and using applications.

Data Science, Instant messaging, Fingerprint on the internet

Resumen

Tanto las redes sociales como los sistemas de mensajería instantánea han crecido en la última década, llegando a ser un medio importante para adquirir y difundir información en diferentes dominios, como negocios, entretenimiento, política, entre otras. Este caso de estudio tiene como objetivo analizar la motivación sobre el aprendizaje de la disciplina Data Science (DS) o Ciencia de Datos, dentro de la población de estudiantes de ingeniería en sistemas computacionales de séptimo semestre del Tecnológico Superior del Occidente del Estado de Hidalgo (ITSOEH), el experimento consistió en aplicar una encuesta en dos momentos, al inicio y final del curso Introducción a la Ciencia de Datos dentro de una escala de Likert, esto para medir el grado de motivación de los alumnos ante el área de DS. Primero se obtuvieron diferentes conversaciones de WhatsApp para posteriormente realizar un análisis por medio del lenguaje estadístico R y el entorno de desarrollo RStudio, se examinaron los datos mediante gráficas para visualizar el comportamiento de los participantes de la conversación. Los resultados del caso de estudio muestran que los alumnos conocen el poder de los datos para las grandes empresas, pero desconocen el uso de los datos generados por nuestra huella digital ante la navegación de internet y uso de aplicaciones.

Ciencia de Datos, Mensajería instantánea, Huella digital en internet

Citation: NAVARRERE-ARIAS, Dulce J., HERNÁNDEZ-GARCÍA, Héctor Daniel and PÉREZ-BAUTISTA, Mario. WhatsApp conversation analysis as a means of motivation for teaching in data science. Journal-Mathematical and Quantitative Methods. 2021. 5-9: 18-24

* Correspondence to Author (email: dnavarrete@itsoeh.edu.mx)

† Researcher contributing first author

Introduction

Data science (DS) is an interdisciplinary field that is responsible for the extraction of knowledge in data. This discipline is supported by Big Data, which consists of storing, processing and analyzing large volumes of data with computer science technologies. Data science is changing the traditional ways of doing things in business, health, politics, education, and in a myriad of fields (Mayer-Schonberger & Cukier, 2013). The objective that data science achieves is in the search for models that describe patterns and behaviors from the data in order to make decisions or make predictions.

Data science feeds on various data as inputs, either from networks of sensors and devices, or from the activity of users on the internet (Batty et al., 2012). Much of this data provides personal information about people's daily activities. People leave a digital trail, either voluntarily or involuntarily when we carry out activities on the internet or on our devices.

We leave our fingerprint when using our mobile phone to pay for services, such as electricity, water, connectivity service, telephone rental and a myriad of processes that we carry out such as online purchases, page views and even the places we visit, thus leaving a mark of personal activity (Calabrese et al., 2011). These data constitute an important source of raw material for the study of human behavior. We are experiencing an important data revolution, in which data acquires increasing value for companies (Marr, 2017).

Given the existence of different social networks as data providers in an indirect way, there is an open field to design data science courses that allow students to experience this data with the languages and tools used for real-world situations so that understand the importance of their digital footprint on the internet (Alonso, López, Font, & Manrique, 2010).

This work presents the practice carried out for the analysis of conversations made by the same students on the WhatsApp platform, implementing languages and tools for data science, and the experimental learning model, where the student learns through experimentation instead of being just receivers. liabilities of knowledge.

Much of the development of Experiential Learning theory (EL – Experiential learning) in recent years they have gained relevance due to the research of David A. Kolb (Kolb, 1983), where you specify the principles of your learning proposal. Exploring the explanatory power and utility of his theory in various disciplines and professions. In the scope of this case study, the Kolb cycle is applied to the field of study of data science (Jacobson & Ruddy, 2004).

According to Kolb, the experiential learning cycle involves four phases:

1. **Concrete Experience Capacity (CE):** be able to engage fully, openly and without prejudice in new experiences.
2. **Reflective Observation Capacity (OR):** be able to reflect on these experiences and observe them from multiple perspectives.
3. **Capacity for Abstract Conceptualization (AC):** be able to create new concepts and integrate your observations into logically sound theories.
4. **Active Experimentation Capacity (EA):** be able to use these theories to make decisions and solve problems.



Figure 1 Learning cycle through experience or Kolb's learning cycle

Reference source: <https://www.actualidadenpsicologia.com/la-teoria-de-los-estilos-de-aprendizaje-de-kolb/>

The following section details the application of each of the phases of the cycle in the developed practice.

Methodology

In the first instance, in order for the students to understand the R language and the steps of experience-based learning, exercises that have no relation to the world around them were replicated in order to control the results.

For the practice, the participation of 19 students from the educational program of computer systems engineering in the specialty of data science was given the web page "Analysis of chats in WhatsApp: Part 1 - Text analysis and Data visualization with R ", with the aim of understanding and implementing the published code having as a data source their own conversations on the platform and thus obtaining some graphs in order to be able to be analyzed in order to understand the behaviors of individuals.

Description of Practice

Taking into account that nowadays any device that collects data, these can be measured with the appropriate knowledge and tools, the practice consisted of identifying and selecting the conversations to be analyzed to transfer them from the mobile phone to your personal computer. Subsequently, a reading and preparation of the data was carried out through the R language, it should be noted that for this practice messages sent in the years 2020 and 2021 were analyzed, segmenting them by seasons of the year.

Once the data has been prepared and segmented, the following types of graphs to be obtained were defined:

1. Frequency of daily messages.
2. Frequency of messages by day of the week.
3. Frequency of messages by time of day.
4. Who has sent the most messages?
5. What are the most used emojis in chat?
6. Most used emojis in chat, by user
7. What are the most used words in the chat?
8. Most used words in chat, by user

In order for the students to have meaningful learning from practice, in addition to being able to recreate a real situation, the phases of Kolb's experimental cycle were applied as follows:

1. **Concrete Experience Capacity (CE):**
The students identified personal WhatsApp conversations that they wanted to analyze between 2020 and 2021.
2. **Reflective Observation Capacity (OR):**
The students implemented code in the R language to visualize findings in their conversations.
3. **Capacity for Abstract Conceptualization (AC):**
The students looked for a way to execute different personal conversations through WhatsApp instant messaging to obtain defined graphs that describe the behavior of the actors in the conversation.
4. **Active Experimentation Capacity (EA):** The students concluded by finding patterns in the conversations; taking into account the graphs obtained, moods, romantic relationships, type of communication between people, icons and words most used by people in the conversation.

Once the methodology for learning under experience and the practice to be carried out has been defined, we take the case study on the motivation that this practice leaves in students to study the specialty in data science, for this a survey was carried out that It was applied to the 19 students in two stages: before carrying out the practice and after having carried it out. The foregoing in order to measure interest in the specialty.

Results

From the practice, the students obtained 8 graphs from the personal conversations extracted from WhatsApp, in which the following was shown.

1. Frequency of daily messages

The students determined that there are some months in which they have more communication, in this case due to the relationship that people have, they conclude that in the months of August and September there was more communication derived from the beginning of classes, however, in summer there is activity between the people involved, see figure 2.

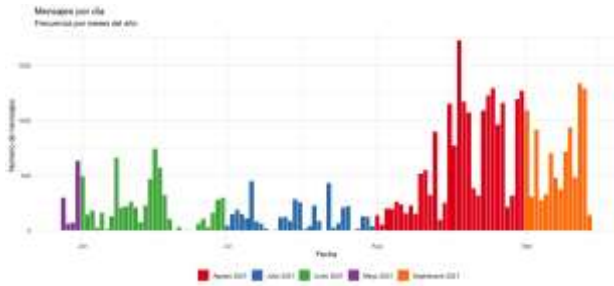


Figure 2. Messages per day, dividing the data by season of the year
 Source of consultation: Own Elaboration

2. Frequency of messages by day of the week

In personal relationships there are behavioral behaviors to follow, it is very common for these behaviors to only be noticed in person, but at the time of having a written communication between two or more people we can analyze the constancy of communication and determine behavior patterns, for example, in figure 3 it can be seen that the interaction of people with days, Saturdays, Sundays and Thursdays.



Figure 3 Frequency of messages per day, dividing the data by season of the year
 Source of consultation: Own Elaboration

3. Frequency of messages by time of day

As I have seen from a conversation, we can determine how frequent the daily messages are, but it can also be determined at what time of day this interaction takes place. In figure 4 it is possible to observe the communication occurs between 10:00 and 11:00.

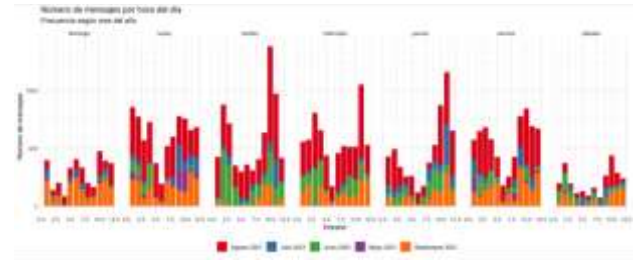


Figure 4 Number of messages per hour of the day. Frequency according to the season of the year
 Source of consultation: Own Elaboration

4. Who has sent the most messages?

In this section you could give an indication about the personality of the individuals involved in the conversation. We could intuit who is more impatient of the two. As can be seen in graph 5, person 1 is the one who sends the most messages to person 2.



Figure 5 Total number of messages per user, dividing the data by season of the year
 Source of consultation: Own Elaboration

5. What are the most used emojis in chat?

Now we have in figure 6 which have been the most used emojis in chat by our users, remember that in these times an emoji says more than a thousand words.

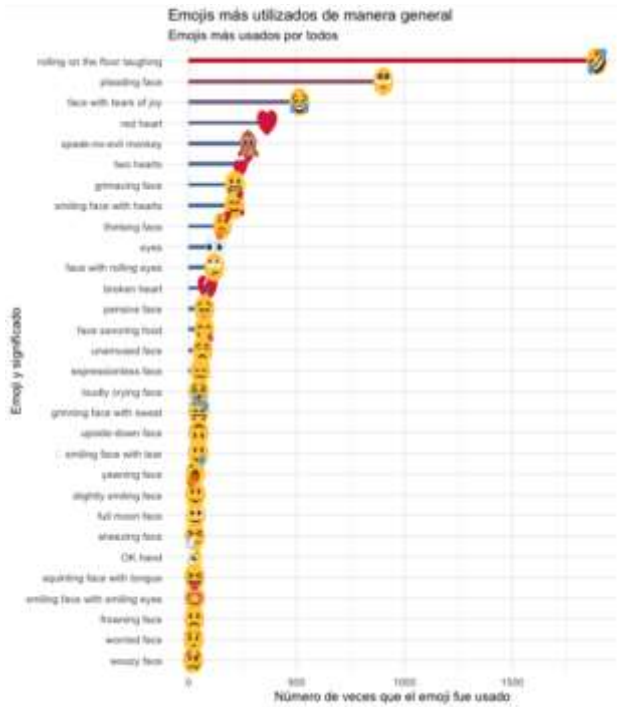


Figure 6 Emojis most used during conversation
 Source of consultation: Own Elaboration

6. Most used emojis in chat, by user

Now figure 7 shows the top 8 emojis most used by user.



Figure 7 Most used emojis in conversation per user
 Source of consultation: Own Elaboration

7. What are the most used words in the chat?

As was done with emojis, we can analyze the frequency of words used in the conversation. In this part, a classification of words was carried out, all those that were wanted to discriminate were added to a set of words, since they are not relevant, such as articles, pronouns, etc. In figure 8 we can see the type of language used by individuals and determine that it is an informal language, as it occurs in a conversation with friends.

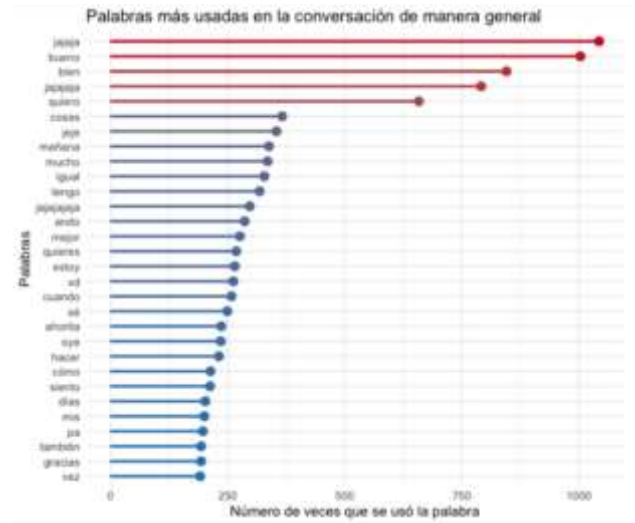


Figure 8 Words most used in conversation in a general way
 Source of consultation: Own Elaboration

8. Most used words in chat, by user

In this part we could say that regardless of the relationship between two people, the richness of each person's language can be determined, as we see in figure 9, the variation in vocabulary is not variable between the two people.



Figure 9 Words most used by user in the conversation
 Source of consultation: Own Elaboration

Now in order to know the perception of the specialty students of Data Science in the 7th semester of the Systems Engineering career about the motivation to continue learning in the area of Data Science, a survey was used that, as shown He explained, it was applied at the beginning and at the end of the implementation of the practice, it allowed to evaluate the impact of exposing students with minimal knowledge of the Data Science area to the interaction with activities related to this area.

The items used in the survey are related to the aspects of motivation and satisfaction of the students around working with tools, technology, techniques and theories related to the area of Data Science. The survey was implemented in a Google form, which consisted of 10 multiple-choice questions on the Likert scale (totally agree, even totally disagree).

The case study was applied to 19 students in the subject Introduction to Data Science of the Computer Science Engineering career at ITSOEH. As can be seen in graphs 10 and 11, the students responded positively regarding the use of the tools and techniques of Data Science since it motivates them to learn more about this discipline, and in satisfaction a response was obtained in the same positive way since the knowledge was approached in an attractive way, easy to use and to understand the operation of these tools.

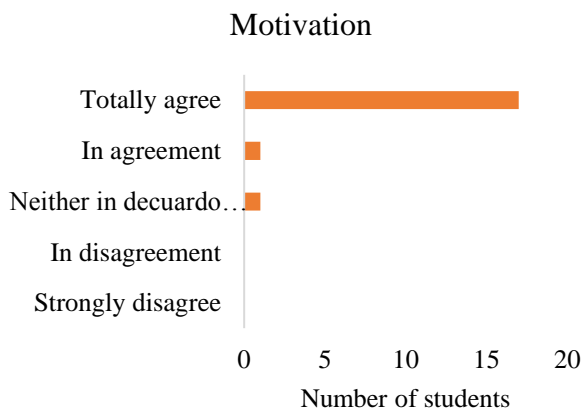


Figure 10 Student's perception of whether the use of tools and techniques typical of Data Science motivates them to learn more about this discipline

Source of consultation: Own elaboration

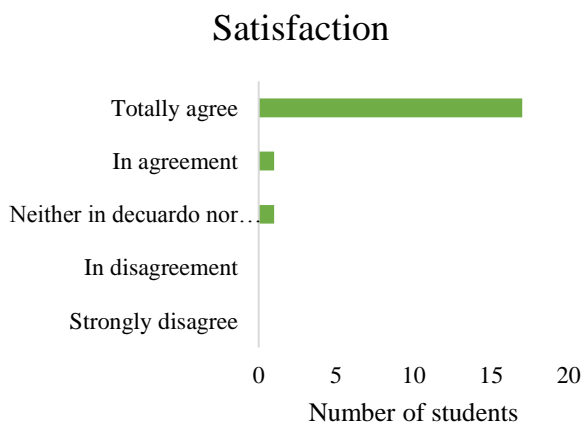


Figure 11 Perception of the student regarding their satisfaction with the techniques of Data Science

Source of consultation: Own Elaboration

Conclusions

Data Science has a wide field of application, the analysis of information shows patterns that we do not know at first glance, being able to represent data through graphs, in this case, describing people's conversations through various data, is representative to understand that data analysis is not only given in information that large companies use. Each element of our fingerprint contributes to this super set of data that allows us to define behavior patterns in society, in this way it is easy to profile users and offer a certain service, product or even to manipulate opinions. In this case study, it was achieved that the students of the specialty of data science, managed to motivate themselves and above all understand that this discipline is not only the task of large companies and that our digital trail together with that of other users is transformed into data valuable, representing large amounts of money.

References

Alonso, F., López, G., Font, J. M., & Manrique, D. (2010). Learner satisfaction when applying an instructional model in e-learning: an experimental study. *Proceedings of the 2nd International Conference on Computer Supported Education - Volume 1: CSEDU*, 141-146.

Batty, M. (2010). «The pulse of the city». *Environment and Planning B: Planning and Design*, 37, 575-577.

Calabrese, F.; Lorenzo, G. D.; Pereira, F. C.; Liu, L. y Ratti, C. (2010). *Analyzing Cell-phone Mobility and Social Events. NetMob-Analysis of Mobile Phone Networks*. Cambridge, MA.

Jacobson, M., & Ruddy, M. (2004). *Open to Outcome: A Practical Guide for Facilitating & Teaching Experiential Reflection*. Wood 'N' Barnes.

Kolb, D. A. (1983). *Experiential Learning: Experience as the Source of Learning and Development*. Prentice Hall.

Marr, Bernard, (2017): *Big data en la práctica: Cómo 45 empresas exitosas han utilizado análisis de big data para ofrecer resultados extraordinarios*. Zaragoza: S.L. TEELL EDITORIAL.

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.