Data Science is a future oil

Abdelhadi, Doaa, BC.

University of people

Abstract

Data science is one of the opened research areas. This paper will focus on data science is power for the future, by collecting data about how people see the future of data science and analyze it by applying data science techniques. Other searches focus on observing the difference between data science and statistic, the history of data science and the future of data science. This paper for each one either works on the business or engineering and medicine. Because it will answer why important to apply data science and how to use data science effectively. The methodology that I used to support the idea of, data science supports itself by itself. I found 66% knowing about data science. 44% would like to become a data scientist or start a data science team. By analyzing text focus on answering the question of using data science effectively. Data science does not have a specific definition. The paper will be supporting the definition of data sciences for all to be easy using it.

Keywords: data science, business, NLP, text analyzing, technology, business, engineering

Data Science is a future oil

Data science is the oil of the future. Governments and companies must be asked to build a Data Science team that is required not an option and everyone asked to have data science skills. Donoho said in your paper called " 50 years data science" published in 2015 "More than 50 years ago, John Tukey called for a reformation of academic statistics. In 'The Future of Data Analysis'" and in 2015 universities started the first program major for data science (Donoho, 2015).

This opens the question of what is Data science that still 50 years until getting in life and spreads extremely quickly across 5 years.

At the same time how the statics see the future of data science "according to International Data Corp. (IDC), worldwide revenue for big data and business analytics (BDA) solutions will reach by $260 billion in 2022, with a compound annual growth rate (CAGR) of 11.9 %". (BISHT, 2019, para.1). We noticed concepts of big data and business analytics in the quotes these concepts and several concepts such as machine learning and self-drive will be covered on the paper. That is a common popular question for everyone decided to study data science.

Data is growwing every day, as reported "over 2.5 Quintillion bytes of data are created every single day, and it’s only going to grow from there. By 2020, it’s estimated that 1.7 MB of data will be created every second for every person on earth" (Ahmad, 2018, para.2).

But the growing data and upper the big data concepts open the question we can't use small data this question will cover in the paper.

We can look at how data science is powerful in medicine "now with genomics data from around 50,000 patients and counting, and 19 participating institutions from around the world, GENIE leaders such as Charles Sawyers, a cancer biologist from the Memorial Sloan Kettering Cancer Center (MSK) in New York City, are struggling to draw actionable insights from the glut of data" (Dana-Farber cancer institute Department of Data Sciences, n.p).

Data science is power almost every industry in our life. For example, how data science is power for the text all of the companies analyze text to understand the behavior of its customers that we can notice from how to the recommends system on Facebook or Amazon more improve than before that. The market has several applications base on data science and machine learning.

The paper will cover common popular question on data science, the top skills required for data scientists and the power of data science for business.

One of the benefits of using data science for business especially here studying the customer behavior and cluster to groups depending on their behavior.when can picked that from the article titled “How to Use Data Science and Machine Learning to Revolutionize 360° Customer Views ”.

The author explanated the role of a data scientist on an organization and how a data scientist can help your company to improve by data that have. I know that when McDonald has explained the statute of MapR company before using data science and it had used the structure data, during applying data science and after applied data science to analyze unstructured data to get insights and add value to the business. And McDonald said, " A data scientist is someone who can draw the lines between the before and after" (McDonald, 2016, p.2).

The author has presented her evidence when she provided us with the result of applied data science. When she said the model created by machine learning is worked well after exploring, cleaning and classifying the browser history data. I know that from the last paragraph "For full-service restaurants (the yellow line), the probability of clicking tracks well with the click-through rate. At a low probability, the click-through rate was 30%, and at a high probability, the model found populations with a 70% click-through rate. So, the model really responded to the method of finding similar customer profiles using their Internet browsing history" (McDonald, 2016, p.5).

Totally, I agree with McDonald's because we can observe the importance of data science and machine learning for business and other areas such as education and how it can help them to add more value. For example, there is much research work on improving the best model to predict if a person has a disease or no. In the last event happening right now "Corona Virus " the World

Health Organization collected data across the countries that the Virus appeared in it to get insights and predict how the Virus will be spread.

Form all the references that I read it, I notice the authors focus on data science is a future oil. We can observe that when McKendrick said AI and machine learning moves from the research and enter in the industry (McKendrick, 2019). He mentions companies turn into using it's data and automation when he said in the last paragraph, companies turn to replace humans by robots. At the same time, Gorbunov said data is future oil to support the future by making discoveries. And he said data science isn't taken us to the answer directly but it is provided us with insights that help to find the answer (Gorbunov, 2020). On another hand, McKendrick mentions data science has tools, algorithms, and methods to process and extract knowledge from structured and unstructured data (McKendrick, 2019). Gibson and others refer to an important point that researchers try to work and build their new knowledge from natural resources because these recourses are lower cost and more efficient (Gibson, 2019). A data science team is already working on natural resources such as on texts that present on social media like Facebook and Twitter to study human behavior to improve their services at the same time, social scientist can use it to understand and build new models to human behavior theory. Facebook on the last policy update is made this data public for scientists to use it. That is open a question why Facebook make data is available for researcher.

In the next section I will be focus on the methodology and findings from the survey I shared.

Methodology

Before I start explorer why I use the quantitative methodology and what are the results which I get it. I picked using the quantitative methodology and I created seven questions on platform Googles formss and share with my friends.

Questions

1. What do you do?
2. What is your work field?
3. Did you hear about a new science called data science?
4. What do you know about data science?
5. Do you use Intelligent systems such as recommendation systems on Facebook (pages, friends), google translations and Etc and you feel it is useful?
6. How did you see it will be useful or useless for your work?
7. Do you plan to (become a data scientist) (start data science team on your Campany?

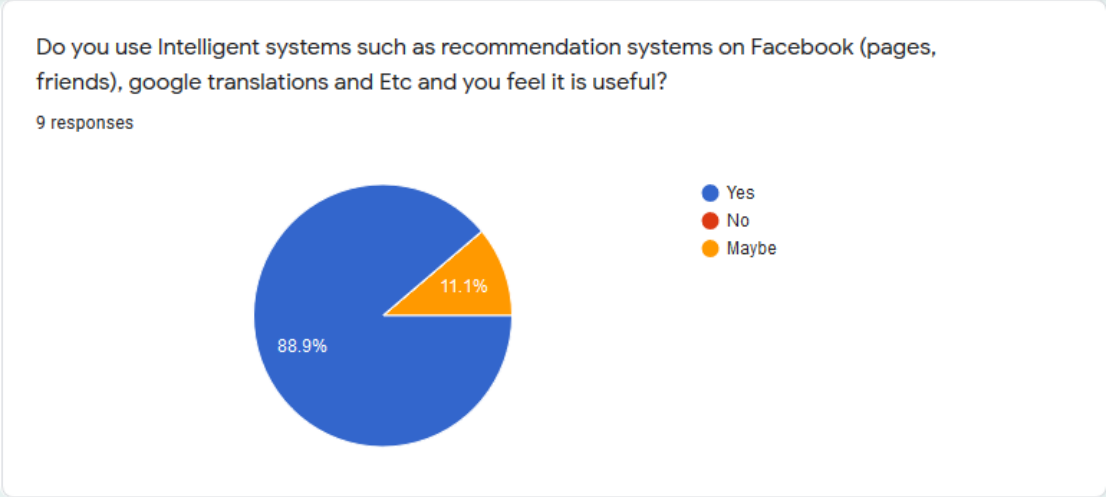
For more information about the design of the online survey, you can see them from (Google_form).

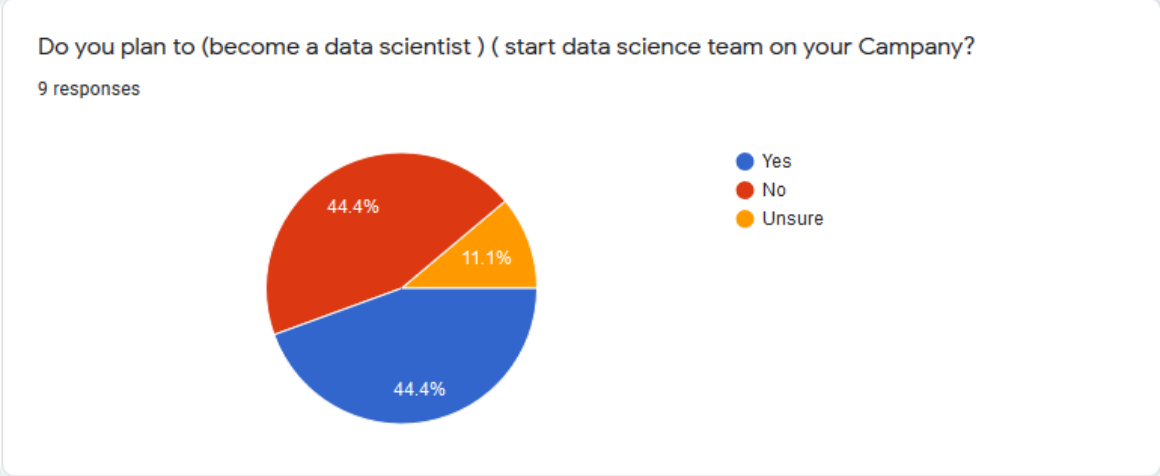
Findings on graphs.

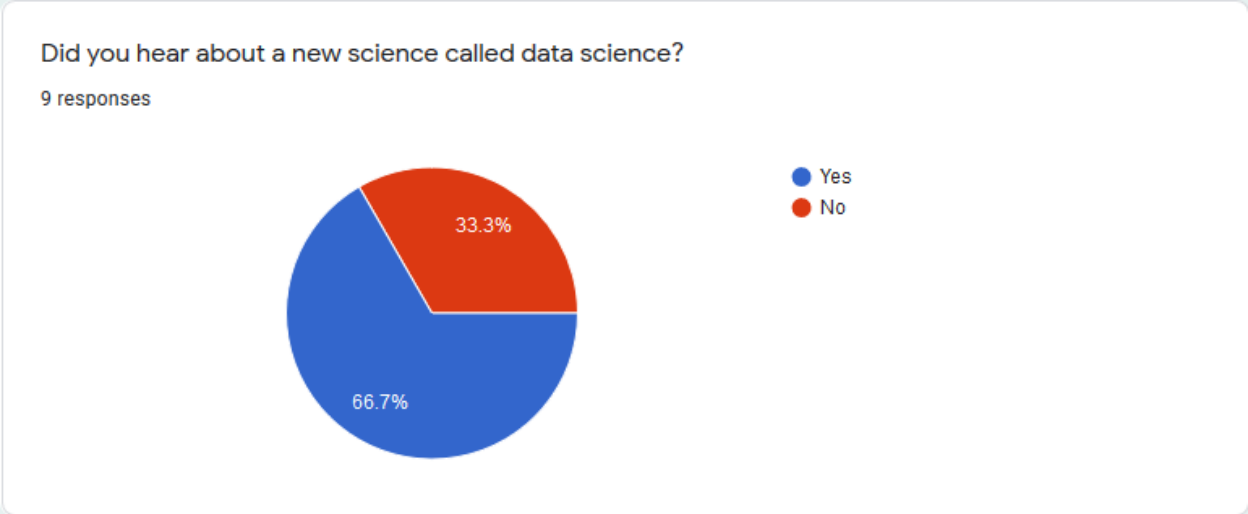
After I explore that methodology which I pick it the time out to explain why I use the quantitative methodology. I used survey research to get information about what people know about data science. Why I picked the survey method because I can share it online.

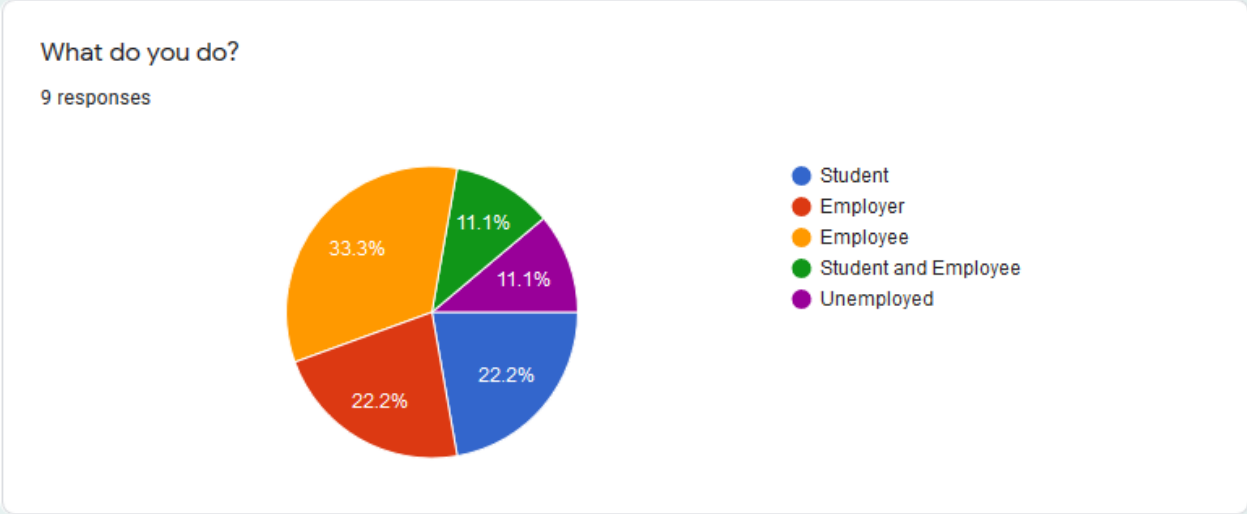
It gives a samples random people does not focus on species sample, more objective, and reducing a complex problem to a limited number of variables. On the other hand, I use statistics to generalize my findings and I can use Google sheet to analyze the results easily, in the future, I can analyze the text which the partitions said.

I would like to mention some points before I explain my findings first, I just received nine responses until now. I build my results depend on it. Second, I have attached some charts that help me to explain my results.









Findings.

Two registered as employers, one of the employers has heard about data science and another one hasn't heard but both of them used intelligent systems and don't plan to become a data scientist. Three registered as employees, two of them have heard about data science and both of them used intelligent systems. one of them, his company has a data scientist team. Other people didn't hear about data science, used intelligent systems and does plan to become a data scientist. Two registered as Student, both of them have heard about data scient, used intelligent systems and they are planning to become a data

scientist. One person registered as unemployed, he\she has heard about data scient, used intelligent systems and he\ she is planning to become a data scientist.

In conclusion, there is the equation between that is believed data science is important and plans to become a data scientist or build a data scientist team.

Conclusion

Form all the references that I read it, I notice the authors focus on data science is a future oil. My findings need more data to accept my thesis or not accept it. Because the data which I collected it is not enough. One of the better chances to support my thesis data that collected about Corona virus and how will use it to help the world. It’s chance to prove data science is a future oil.

References

Ahmad.I. (2018, June 15). How much data is generated every minute? [Infographic].*Social media today*.Retrieved from https://www.socialmediatoday.com/news/how-much-data-is-generated-every-minute-infographic-1/525692/

BISHT.P.(2019, May 23). How will Data Science and Data Scientists in 2030 be like? *Houseofbots.* Retrieved from *https://www.houseofbots.com/news-detail/12029-1-how-will-data-science-and-data-scientists-in-2030-be-like*

David.D. (2015, Sep 18).50 years of Data Science. Retrieved from http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf

Dana-Farber cancer institute Department of Data Sciences. (n.d.).Big data takes on cancer.*Natural.*Retrieved from https://www.nature.com/articles/d42473-019-00035-5

Gibson, K. J., Streich, M. K., Topping, T. S., & Stunz, G. W. (2019). Utility of citizen science data: A case study in land-based shark fishing. *PLoS ONE*, *14*(12), e0226782. Retrieved from https://link.gale.com/apps/doc/A612482845/STND?u=lirn17237&sid=STND&xid=4da5d8c3

Gorbounov, S. (2020, January 28). Making room for data science in the humanities. *UWIRE Text*, p. 1. Retrieved from https://link.gale.com/apps/doc/A609224120/AONE?u=lirn17237&sid=AONE&xid=b238f99f

McDonald.C. (2016 August 08). How to Use Data Science and Machine Learning to

Revolutionize 360° Customer Views. Retrieved from https://mapr.com/blog/how-use-data-science-and-machine-learning-revolutionize-360-customer-views/

McKendrick, J. (2019, December-January). What's Ahead in Data for 2020--And the Coming Decade. *Database Trends & Applications*, *33*(6), 4+. Retrieved from https://link.gale.com/apps/doc/A611172204/CDB?u=lirn17237&sid=CDB&xid=4ba801d1