A social media approach for improving decision-making systems

Islam SADAT
Computer science department.
University of science and technology Houari Boumediene
Algiers, Algeria
islam_sadat@outlook.fr

Kamel BOUKHALFA
Computer science department
University of science and technology Houari Boumediene
Algiers, Algeria
boukhalk@gmail.com

***Abstract*—The appearance of social networks has revolutionized the way of interaction and exchange of information on the internet, and has led to an explosion of data on these networks. The availability of this data via a set of APIs triggered a wave of initiatives and work aiming to make use of this massive amount of information in order to extract knowledge that will facilitate the decision-making process. In fact, due to growing business competition, leaders from different sectors have already started using reporting tools, making it possible to provide reliable and relevant analyses at the right time. In this paper we present an approach for improving an existing decision-making system by combining its core datawarehouse with a social one. This social datawarehouse built using extracted tweets will expand the analyzes possibilities of the existing system, and will allow carrying out various analyses, facilitating the decision-making process, and this for any sector.**

***Index Terms*—Datawarehousing, Social Media, Decision Making, Data Analysis**

I. Introduction

Contemporary information and decision support systems have been essential to the proper functioning and growth of successful businesses around the world for more than two decades. Data warehousing and OLAP technology are at the center of these systems and have been instrumental in analyzing data in multiple areas such as manufacturing, retail, transport, health care, education, research and government.The data warehousing technology as well as its underlying techniques have been extended to provide better performance, by taking advantage of the emergence of new data types and sources, especially data of the public shared on the phenomena of social media. Indeed, social media have shaped the last two decades of the 21st century, and is considered as a revolution that sparked almost all ways of life. The users content on these sites represents huge volumes of data that is generated at a high rate and attracts a lot of research and interest. Since then, companies and organizations with decision-making systems have sought to make this continuous flow of information concerning them beneficial, and to use it as an asset to facilitate decision-making. Table I represents the number of users by social network for the year 2020.

Datawarehouses as well as multidimensional analyzes are essential for any decision-making process; however, the absence of tools allowing the use of social networks in decision support constitutes a niche research field that aroused the interest of a good number of researchers, what resulted in various works aiming to carry out multidimensional analyzes on social data. We aim through this study to improve an existing decision-making system by transforming its datawarehouse into a more robust one, composed of its original dimensions and aggregations, as well as new social activity tables and dimensions built using social media data (Twitter). This hybrid datawarehouse will offer the possibility of performing more consistent analyses, and will facilitate the decision-making process. In order to build this system we encountered the following challenges :

TABLE I
Number of users by social network

$Facebook$	2 740 000
$Youtube$	2 291 000
$Instagram$	1 221 000
$Twitter$	353 000
$Quora$	300 000

- The extracted social data (tweets) is multidimensional, time stamped, geographical and composed of many attributes. Designing a database architecture capable of storing this type of social data was a challenging subject.
- Choosing the most appropriate social dimensions that allow the fusion of social and non-social data.
- Summing the social data and finding the correct aggregations with their respective dimensions was a major step in order to answer the users queries in the most efficient way.

In this paper, we first explain the sources of the social data, the extraction process as well as the preparation of the social corpus. Then we introduce the architecture of the activity and dimensions tables which will store the aggregated social information. Next we discuss the different multidimensional analyses possibilities offered by our system. Finally, we conclude the paper and discuss future research.

II. Literature Background

A. Datawarehouse

A data warehouse is a type of data management system designed to enable and support business intelligence activities, particularly analytics [18]. Data warehouses are intended only for performing queries and analyzes. They often contain large amounts of historical data. The data in a data warehouse usually comes from a wide variety of sources, such as application log files and transaction applications. According to the firm Gartner [19], the business intelligence market reached a worldwide turnover of 27 billion dollars in 2019. Business intelligence is targeting to improve business decision-making on the basis of established facts and offers non-IT decision-makers a transversal vision of strategic information.

B. OLAP Technology

OLAP (online analytical processing) is the application of complex queries to large amounts of historical data, aggregated from OLTP databases and other sources, for data mining, analysis and business projects intelligence. In an OLAP database, data is integrated through a process called extract, transform, load (ETL). This pre-aggregation is what dramatically reduces response time to complex queries. Each query involves one or more columns of data aggregated from multiple rows. Examples include year-over-year financial performance or lead generation marketing trends [21]. OLAP databases and data warehouses allow analysts and decision makers to use custom reporting tools to turn data into information. Figure 1 shows an example of a multidimensional cube.

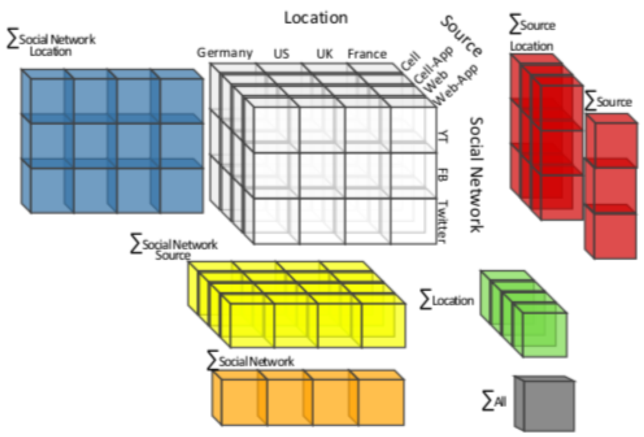

Fig. 1. A three-dimensional cube (gray) with its projections (colored)

C. Decision-making system

A BI (business intelligence or decision-making system) is designed to answer questions like "What happened?", "Why did it happen?" And "What's going to happen?" and this with the aim of enabling organizations to understand their internal and external environment, through the acquisition, collation, analysis, interpretation and systematic use of information [22]. A BI system typically performs customer profiling, customer support, market research, market segmentation, product profitability, and inventory and distribution analysis. These functions can be accomplished through data warehousing, OLAP, data mining, descriptive and predictive analytics. BI systems have played a key role in running a successful business, and today it's hard to find a successful business that hasn't taken advantage of BI technology. Figure 1.1 shows the process of entering and leaving data in a BI framework.

D. Twitter

Microblogging is a network service with which users can share messages, links to external websites, images, or videos that are visible to users subscribed to the service. Messages that are posted on microblogs are short in contrast to traditional blogs. Like It or Not: A Survey of Twitter Sentiment Analysis Methods 28:5 Currently, a number of different microblogging platforms are available, including Twitter, Tumblr, FourSquare, Google+ and LinkedIn. One of the most popular microblogs is Twitter, which was launched in 2006 and since then has attracted a large number of users. Currently, Twitter has 284 million users who post 500 million messages per day. Due to the fact that it provides an easy way to access and download published posts, Twitter is considered one of the largest datasets of user generated content[3]

E. Big Data

The rapid development of the Internet of Things (IoT) results in a massive explosion of data generated from ubiquitous wearable devices and sensors [5]. The unprecedented increase of data volumes associated with advances of analytic techniques empowered from AI has led to the emergence of a big data era [5], [6]. Big data has been employed in a wide range of industrial application domains, including healthcare where electronic healthcare records (EHRs) are exploited by using intelligent analytics for facilitating medical services. For example, health big data potentially supports patient health analysis, diagnosis assistance, and drug manufacturing [7]. Big data can be generated from a number of sources which may include online social graphs, mobile devices (i.e. smartphones), IoT devices (i.e. sensors), and public data [8] in various formats such as text or video. In the context of COVID-19, big data refers to the patient care data such as physician notes, X-Ray reports, case history, list of doctors and nurses, and information of outbreak areas. In general, big data is the information asset characterized by such a high volume, velocity and variety to acquire specific technology and analytical methods for its transformation into useful information to serve the end users.

F. Sentiment Analysis

Sentiment analysis is a natural language processing techniques to quantify an expressed opinion or sentiment within a selection of tweets . There are two main approaches for extracting sentiment automatically which are the lexicon-based approach and machine-learning-based approach [9] The sentiment can be found in comments or tweets to provide useful indicators for many different purposes, and its mainly categorized into 3 groups: negative, positive and neutral sentiment words.

III. Related work

Many researchers are working on analysing data on Twitter social media, some key contributions provide support for finding users behaviors and situations in the different cases while happening around the world, for this some of the essential papers are included in this section.

T. Yigitcanlar and al[9], proposed a social media data analyzing study, by collecting 100k tweets generated by Australian users only, in order to capture the attitudes and perceptions of the Australian community during the covid-19 pandemic. The main objective of their study was to exploit social media in guiding interventions and decisions of the authorities, and to identify community needs and demands in a pandemic situation.

K. H. Manguri and al[10], been pulled out using python programming language Tweepy library a dataset of Tweets. The extraction of the tweets was based on two specific hashtag keywords, which are ("COVID-19, coronavirus"). The date of searching data is seven days from 09-04-2020 to 15-04-2020. Then they used python TextBlob library for analysing the users emitted sentiments. The obtained measures were after that represented graphically.

Vijay and al[11], gathered tweets regarding COVID-19 from November, 2019 to May, 2020 in India. Multiple datasets were created Month-wise, then combined to analyze the people's reactions towards Lockdown in June, 2020 and about everything related to COVID-19. The general feeling was negative at first shifted towards positive and neutral comments. In April, 2020 most comments were Positive and about winning against Corona virus.

R. Lamsal[12], prsented a large-scale Twitter dataset with more than 310 million COVID-19 specific English language tweets and their sentiment score, as well as GeoCOV19Tweets, The dataset's geo version. Lamsal's paper discussed the datasets design in detail, and the tweets in both the datasets were analyzed, giving a better understanding of spatial and temporal dimensions of the public shared tweets related to the ongoing pandemic.

Ben Kraiem and al[22] presented a multidimensional modeling tool. Contrary to the work cited above, the tool named OLAP4tweets exploits the association of OLAP technology with data mining techniques to allow an analysis centered on the aggregation of metadata on Twitter users and their web activity. messaging.

In [23] the authors present a tool called SocialCube. The latter helps organize social media data into multiple dimensions and hierarchies for efficient viewing and visualization of information from multiple perspectives, through the application of multidimensional manipulations of social data cubes into different dimensions.

The research papers cited previously encounter many limits. Most of these studies expressed awareness of the importance of using social data. Some of them created informative visuals using a social corpus, but didn't explore the multidimensional aspect of the social data. The others built datawarehouses with data from twitter, allowing multidimensional analyses, but did not make use of already non social datawarehouses. Our study exceeds the limits of other studies by mixing social and non social data in a single hybrid datawarehouse, meaning that our contribution covers the user of social media, it's multidimentional aspect improves the decision making process.

IV. Data and Methodology

The process of improving the performances of a decision making system is presented in this section. The hybrid datawarehouse resulting in our approach is mainly composed of two parts:

1) A non social part (activity and dimension tables): It represents the datawarehouse behind the decision-making system concerned by the social aspect improvements. We have chosen a covid-19 decision-making system as a study case. Its datawarehouse is composed of activity tables and dimension related to the pandemic, allowing visualisation of the variation of number of deaths, recoveries and cases due to covid-19.
2) A social part (activity and dimension tables): This part represent the new activitiy and dimension tables established using attributes and aggregations from a twitter social corpus set up beforehand. This social corpus is composed of an important number of social data(tweets), treated, analyzed and enriched.

The merging of these two parts will result in what we named a "hybrid datawarehouse". We will in what follows explain the steps for setting up the two parts, as well as the fusion process.

A. Non social datawarehouse part

In our case of study we will use "covid.sql", a file comprising a multidimensional cube having as a core unit aggregations related to coronavirus, and that allow to follow the evolution of this pandemic. This cube has two dimensions that are time and location. We have chosen this issue as a case study given the impact this pandemic had globally and the availability of information on social networks, more specifically on TWITTER. As shown in 2, the covid-19 datawarehouse is composed of a single activiy tables with 3 aggregates :

- **number of deaths**
- **number of recoveries**
- **number of cases**

This activity table is related to two dimension tables : **Location** and **Time**, which will allow temporal and geographic analyses. It is clear that the scope of analysis of this datawarehouse is very limited, and we will show in what follows how to extend this scope by setting up a datawarehouse of social data related to coronavirus.

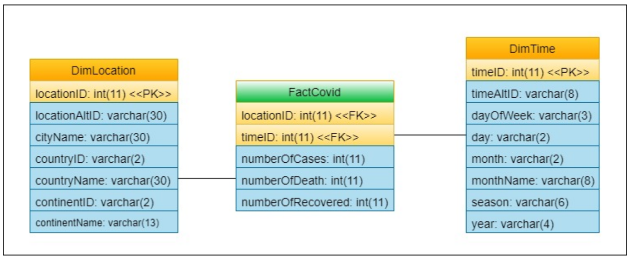

Fig. 2. Schema of the covid-19 datawarehouse (non social datawarehouse)

B. Social datawarehouse part

This part represents the added value that our approach brings to the decision-making system. This social part is built thanks to a social corpus composed of tweets mentioning the coronavirus. We will detail in what follows the stages of construction of the corpus, as well as the definition of the architecture of the social darawarehouse

1) Building the social corpus: Twitter social media network provides across Twitter Developer Platform a set of free APIs that allow retrieval of tweets· The most used API is the Twitter REST API. It basically takes key words as input in order to extract relevant tweets. After trying different open source tools, we concluded that the python library ”Tweepy” is the most convenient one for extracting data from Twitter API. ”Tweepy” offers the possibility of fetching tweets by location, hashtags, keywords and date. In spite of that, the built dataset for this study is general and doesn’t concern a specific country or continent, considering that coronavirus is a world wide pandemic. Despite of having many limitations from Twitter API, 126,000 tweets relative to coronavirus were successfully withdrawn from 29/02/2020 to 31/05/2020. The gathered data was at first saved as csv files. Table II shows indicators relative to the dataset.

TABLE II
COLLECTED DATA INDICATORS

Indicator	*Value*
No. of tweets	126,000
No. of continents	6
No. of countries	159
No. of languages	52

The text of tweets very often contains elements that do not interest us in our case study, such as URLs, Hashtags, Mentions, Emojis, Smileys, JSON and even words specific to Twitter, such as ”RT” for example which means ”Retweet”. Therefore, text filtering is extremely important to keep only representative content. It is essential to enrich the data extracted in the previous step, and this in order to offer different possibilities of analysis. Enrichment is done as follows:

- Language detection: The language used in tweets allows to calculate the number of tweets for each language.
- Detection of the country and the continent: Geo-location will for example make it possible to know the global distribution of tweets according to the continent.
- Detection of the feeling expressed: To classify the publications contained in the corpus according to the emotion released, we use a lexical technique of sentiment analysis.

2) Social datawarehouse schema: The production of a social data-warehouse is at the center of our approach, it goes through two phases:

- Choice of the modelization scheme: The modeling scheme we have chosen to model our social data warehouse is the constellation scheme. The constellation diagram is the amalgamation of several star diagrams that use common dimensions. For this, we must model our different datamarts based on the star diagram.
- Dimensional modeling process: After studying the Twitter meta-model as well as the format of a ”tweet”, we found that it is possible to have two processes to model. The activities that we are going to model are: ”Average sentiment of tweets” And ”Distribution of tweets”.
  Figure 3 represents the schema of ”Average Tweets Sentiment” activity.

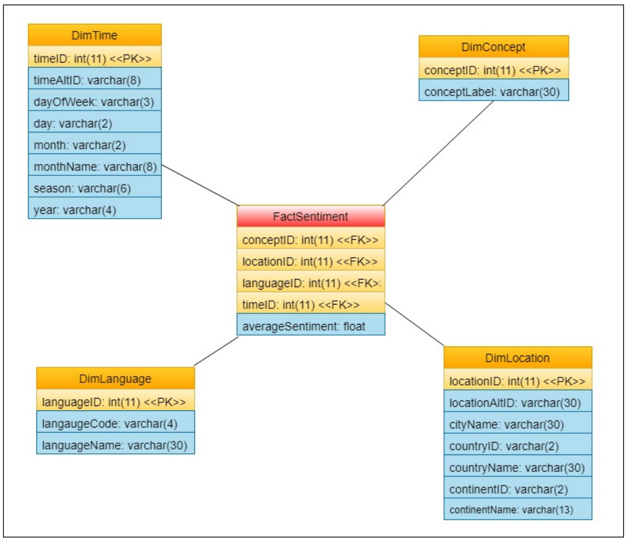

Fig. 3. Star model of the ”Average Tweets Sentiment” activity

Figure 4 represents the schema of ”Distribution of tweets” activity.

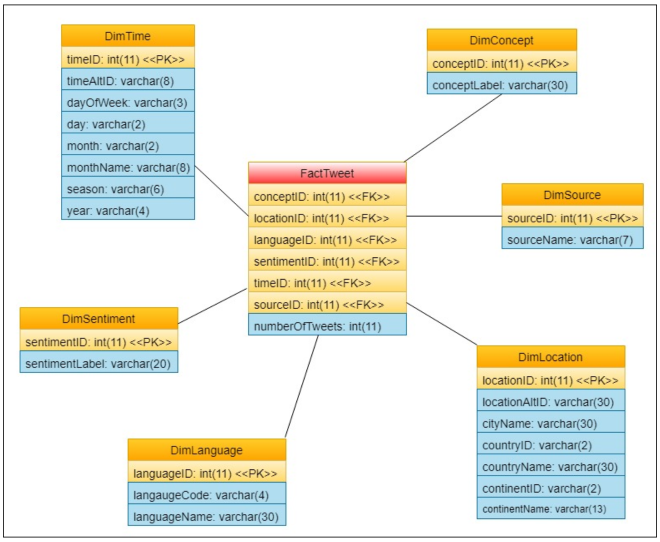

Fig. 4. Star model of the "Distribution of tweets" activity

The two previous schemes are merged into a single constellation schema as shown in figure 5.

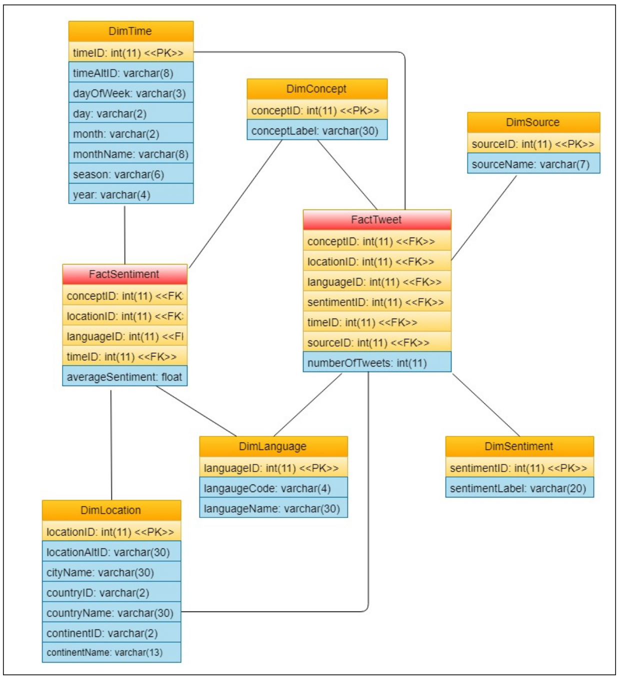

Fig. 5. Constellation scheme of the social datawarehouse

3) Building the hybrid datawarehouse: Merging the two datawarehouses presented above will result in a datawarehouse structured according to a constellation scheme. This datawarehouse will be composed of social and non-social data, and will offer axes of analysis, and the possibility of much richer OLAP operations. The merge between the two datawarehouses is done based of the common dimensions between these two. In our case the shared dimensions are:

- Time
- Location

the hybrid datawarehouse scheme is shown in figure 6.

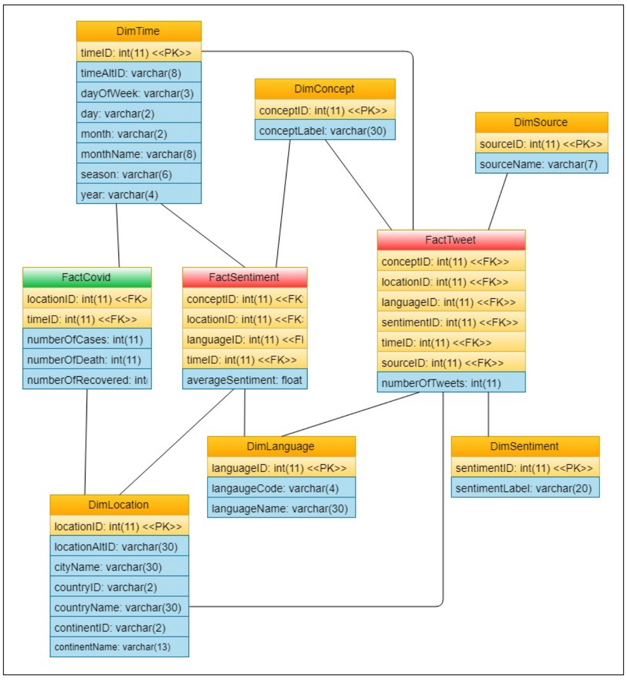

Fig. 6. Constellation scheme of the hybrid datawarehouse

V. MULTIDIMENSIONAL ANALYSIS

It is in this step that the importance of our work is highlighted. Indeed, at this level we run various OLAP queries on the data warehouse modeled by the set of multidimensional cubes. The results of OLAP queries will be made available in the form of graphical visualizations. In this section of the paper we present the different analyses performed of the social coronavirus datawarehouse. The procedure of each analysis will be explained, as well as the obtained results.

A. Geographical analysis

In this phase we analysed the tweets by based on the Location dimension, using the Tweets Distribution cube. Table III show the number of tweets related to coronavirus by continent.
By analysing the report[14] of the ECDC(European Centre for Disease Prevention and Control) regarding the number of deaths due to covid-19 by continent at the end of Mai 2020, we noticed that the top three continents in number of deaths are Europe, America and Asia, the same continents with the more number of tweets relative to coronavirus in our study's dataset.

B. Linguistic analysis

Analyzing in which languages social data is shared can be very beneficial. It may indicate to different entities which languages to use for communicating on social networks in

TABLE III
NO. OF TWEETS BY CONTINENT

Continent	*No. of tweets*
Asia	23438
Europe	18487
North America	10370
South America	8075
Africa	2614
Australia	529
Unknown	62487

order to have great impact and audience. Figure 7 shows the analysis of the distribution of social data relating to coronavirus as a function of language.

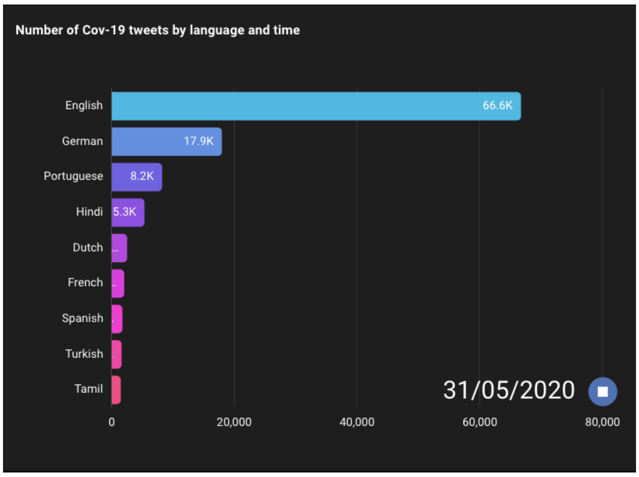

Fig. 7. Distribution of tweets by language

C. Sources type analysis

This analysis consists of distributing the tweets according to the emission sources (Android, iOS, Web). The results of this analysis could prove to be of great interest for choosing the computer software format (Mobile application or website) for having the maximum audience. The numbers of tweets by source are presented in Table IV.

It is clear from the obtained above that more than 70% of tweets were generated from mobile devices, since most internet users today communicate through smartphones.

Figure 8 shows the distribution of social data related to coronavirus by emission source for each continent.

D. Sentiment analysis

This type of analysis allows the visualisation of the variation of sentiments and opinions of the public regarding covid-19 pandemic. The Tweets sentiment activity present in our social datawarehouse is very useful for analysing displaying the number of tweets by sentiment. Figure 9 presents the distribution of social data on "coronavirus" as a function of sentiment.

TABLE IV
NO. OF TWEETS BY SOURCE TYPE

Sourcetype	*No. of tweets*
Android	47196
Iphone	23958
Web	29172
Unknown	25674

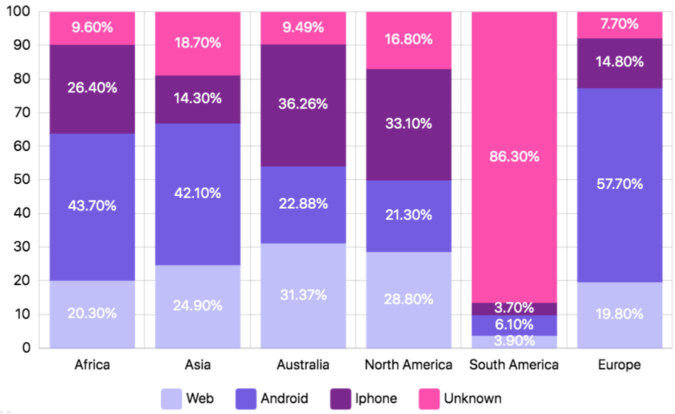

Fig. 8. Distribution of tweets by source for each continent

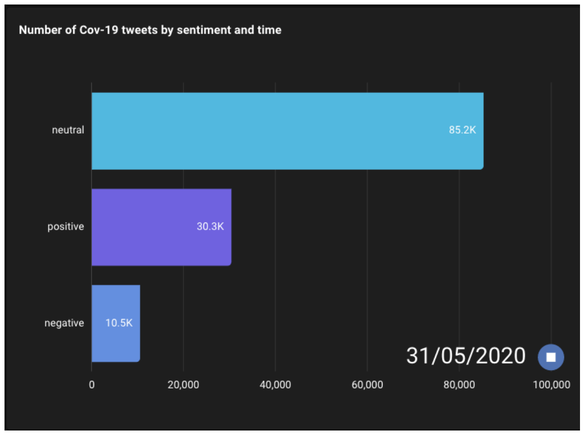

Fig. 9. Distribution of tweets by sentiment

In this study, the bar chart demonstrates emitted sentiment over Twitter starting from February 29th, 2020 to Mai 31st, 2020. It is clear that the total number of tweets is 126K. Overall, more than 30.3k of people published optimistic views, while only around 10.5k of the tweets were negative. However, the No. of neutral tweets was significantly high (85.2k). Such a

Large quantity of neutral tweets might have happened because most of the tweets contained were facts rather than opinions, or because of the presence of a lot of prayers phrases that do not express negative nor positive emotions. The different analyses and indicators presented above were obtained using a tool that we developed as part of our study. In the following section, we will explain in detail the different stages of the implementation of our tool as well as the technologies used.

E. Redundant topics analysis

Although the existence of different NLP techniques for extracting efficient keywords, we have chosen in this work to analyse define relative subjects based on the words redundancy in the tweets texts. This redundant items will most likely offer significant new perceptions, especially after cleaning the data from non significant words like stop words. The more the topic is mentioned, the bigger the node is in the topics cloud. Figure 10 shows the node corresponding to the most redundant topic.

Fig. 10. Most redundant topic node in covid-19 dataset

Table V shows the four most redundant topics in our study data-set and their No. of occurrences.

TABLE V
FOUR MOST REDUNDANT TOPICS IN DATA-SET

Topic	*No. of occurrences*
confirmed cases	5668
corona virus	5173
vote amp	2758
cases deaths	1547

VI. TOOL IMPLEMENTATION

This section is devoted to the detailed description of the implementation of our application which we named ”Cubes Creator”. We will expose the different produced modules as well as the interfaces of our application.

A. Development architecture

In order to provide a tool available on different types of devices and operating systems, we have chosen to develop a web application based on the RESTFUL architecture [38], and this using two very popular technologies which are the VUE.JS framework for the development of the front-end part of the application and the FLASK framework for the back-end part.

B. Developed modules

In what follows, we will detail the different components of our tool by presenting the graphical interfaces relating to the features offered. Our tool essentially consists of (04) modules:

- External data integration module
- Social corpus construction module
- Social data cube construction module
- Social data cube manipulation module

The values of the ”Consumer key”, ”Consumer secret”, ”access token” and ”access secret” necessary for the user of the Twitter API are previously integrated at the application level, so the user does not have to. only have to enter his username and password to authenticate. Figure 11 shows the authentication interface.

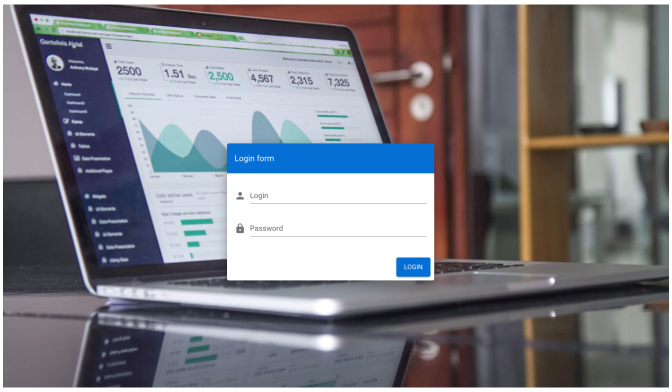

Fig. 11. Authentication interface

C. External data integration module

We allowed the user of our tool to upload external data relative to covid-19 (number of deaths, number of cases). The uploaded information will be be displayed in a relevant dashboard as shown in figure 12 .

D. Social corpus construction module

The construction of the social data corpus is a consistent step in our project. This is the first step in the ETL(Extract-transform-load) process, it aims to extract useful data from specific sources for use in next steps. Our tool offers the possibility of extraction data from the build dataset presented in previous sections, or real time extractions from Twitter using Twitter’s REST API. Figure 13 represents the interface for constructing the social corpus. 13 .

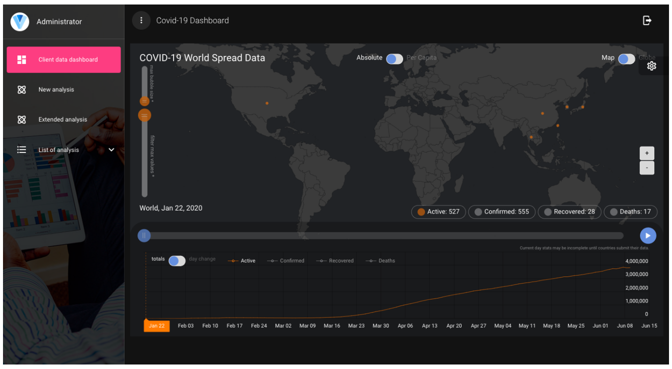

Fig. 12. Global visualization interface of the evolution of the convid-19 pandemic

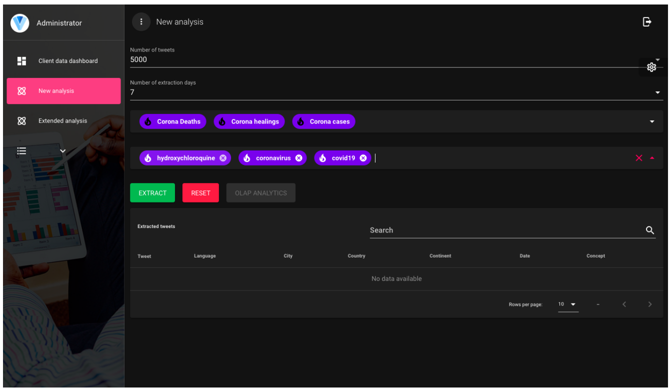

Fig. 13. Social corpus construction interface

E. Social data cube construction module

To apply OLAP technology we calculate from the social corpus the values of the measures according to the dimensions. These calculated values will be inserted at the level of the social part of the datawarehouse, completing the non social part representing in our case of study data related to the covid-19 pandemic.

F. Social data cube manipulation module

Multidimensional cubes produced from the global data warehouse (social data and external data) will allow the user to get a better analysis experience, considering the reduced execution time of OLAP queries, which is due the aggregations made during the creation of the cubes. We will detail in what follows the different analyzes that our application offers.

1) Sentiment analysis of social data: Sentiment analysis is very interesting for decision making. It gives an idea of the opinions expressed by a population concerning a specific subject. Thanks to the multidimensional cube "CubeTweet", we can for example execute an OLAP query manipulating the measure "numberOfT-weets", to know the number of positive, negative or neutral tweets dealing with the subject "coronavirus", and this at a location and at a specific date, as shown in the figure 14.

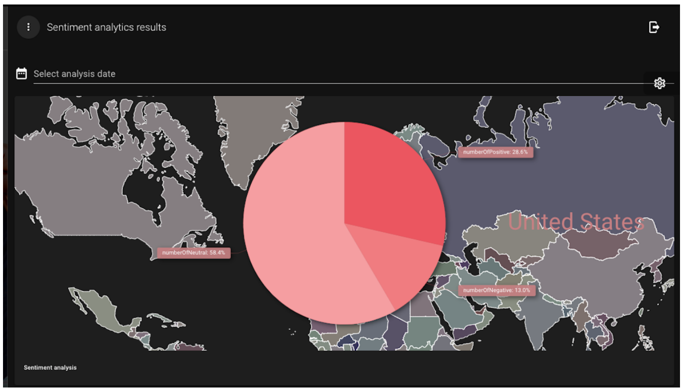

Fig. 14. Analysis of the sentiments of tweets in the USA talking about the "coronavirus" according to time and place

2) Analysis of the distribution of social data: Having multiple dimensions linked to the "CubeTweet" fact table allows us to perform distribution analyzes against multiple axes. That is, get different distributions of social data. Figure 15 represents the interface for visualizing the distribution graphs of social data.

Fig. 15. Social data distribution analysis interface

3) Correlation analysis: This aspect of the application offers the user the possibility of comparing two different graphs, to be able to verify the existence of correlation between external covid-19 data uploaded by user and social data. The following figure represents the graphs of a comparative analysis between the variation of the global sentiment emitted on Twitter concerning the "coronavirus", obtained by exploiting the social cube "CubeSentiment", and the variation of the number of deaths, cures and active cases , obtained by exploiting data provided by the users. Figure 16 shows the interface for comparing the variation in customer metric values with the variation in sentiment on Twitter.

G. Other features of our tool

1) Analysis of related topics: The analysis of relative subjects is the extraction of all the most redundant terms in the social corpus, and which are considered relevant to the concepts on which the social corpus was built. This functionality does not depend on OLAP technology, but we

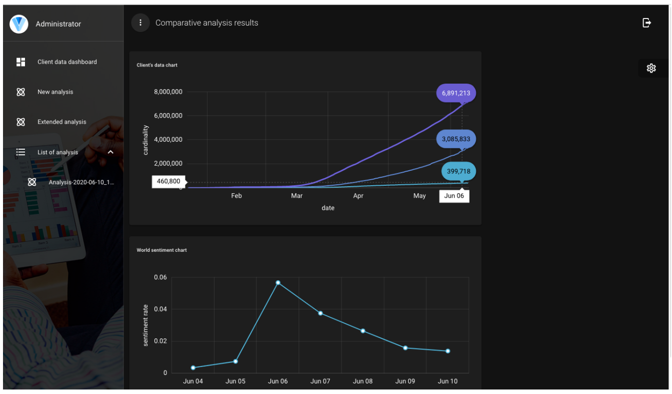

Fig. 16. Analysis interface between the measurements of the user provided data and the social cubes

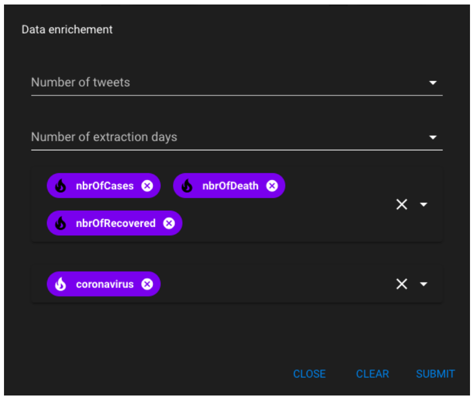

Fig. 18. Data enrichment interface for an analysis

considered it relevant to integrate it, because it offers the user the opportunity to broaden his field of analysis to new aspects. Figure 17 shows the interface for viewing topics related to a concept extracted from social data using our application.

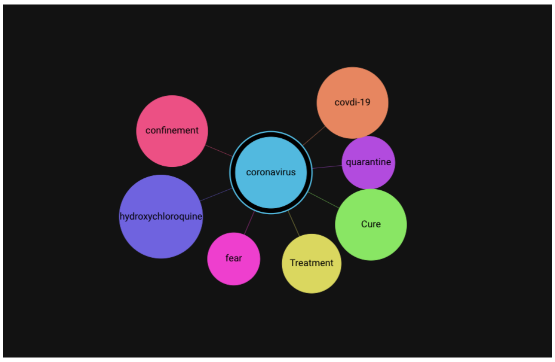

Fig. 17. Analysis interface for related topics extraction

2) Enrichment of an analysis: It is possible at any time for the user to enrich the social corpus of an analysis by performing a new data extraction. The tweets will be processed and added to the set of tweets for the relevant analysis. Figure 18 shows the interface for enriching an analysis.

VII. Conclusion

The huge share of opinions on social media is undoubtedly an important draw for both businesses and world wide organizations. In an increasingly complex and competitive environment, various institutions have become aware of the importance of exploiting this social data, in order to obtain more powerful decision-making and knowledge extraction tools, allowing them to make strategic decisions. timely. If the stakes largely justify this desire, we must not neglect the technical constraints to be overcome. In this modest work, we have focused on proposing an approach to exploit social data in order to improve decision-making systems. More precisely, data from the social network Twitter. We concluded that providing the ability to apply comprehensive analytics to social data is a critical asset in facilitating decision making, and our tool was designed for that purpose. Indeed, it allows you to perform multidimensional OLAP analyzes on data from Twitter, in order to offer the user several axes of analysis, and thus offer him new perspectives not available in a traditional decision-making system.

The tool we have developed allows the user to perform a social data extraction, either directly from the Twitter API for instantaneous analyses, or from a manually prepared corpus of data that can be enriched by the tool at any time. This second possibility is intended for analyzes of liabilities. Subsequently, the previously constructed social corpus will be used to calculate the values of different measures according to several dimensions, which will be saved at the application data warehouse according to a previously defined scheme. Finally, the data warehouse set up will be used to produce multidimensional social data cubes, which our "Cubes Creator" tool will use to perform a set of several OLAP operations.

Our approach is always subject to improvement, so as perspectives of our work, we propose:

- Implementing a sentiment analyzer using machine learning techniques for better classification.
- Developing an android version of the tool given the emergence of smartphones and tablets
- Using a professional Twitter account that can extract an unlimited number of tweets for more consistent analyses.

References

[1] Qiu W, Rutherford S, Mao A, Chu C. The pandemic and its impacts. Health Cult Soc. 2017;9:1–11.

[2] Alsaeedi, Abdullah & Khan, Mohammad. (2019). A Study on Sentiment Analysis Techniques of Twitter Data. International Journal of Advanced Computer Science and Applications. 10. 361-374. 10.14569/IJACSA.2019.0100248.

[3] H.Li, S.-M. Liu, X.-H.Yu,S.-L.Tang, and C.-K. Tang, "Coronavirus disease 2019 (COVID-19): current status and future perspectives," International Journal of Antimicrobial Agents, vol. 55, no. 5, p. 105951, 2020.

[4] V.R, M.Alazab, S.Srinivasan, Q.-V.Pham, S.Kotti Padannayil, and K.Simran, "A visualized botnet detection system based deep learning for the internet of things networks of smart cities," IEEE Transactions on Industry Applications, 2020, in press.

[5] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, "Big data analytics: a survey," Journal of Big data, vol. 2, no. 1, p. 21, 2015.

[6] K. Priyanka and N. Kulennavar, “A survey on big data analytics in health care,” International Journal of Computer Science and Information Technologies, vol. 5, no. 4, pp. 5865–5868, 2014.
[7] M. Cottle, W. Hoover, S. Kanwal, M. Kohn, T. Strome, and N. Treister, “Transforming health care through big data strategies for leveraging big data in the health care industry,” Institute for Health Technology Transformation, http://ihealthtran. com/big-data-in-healthcare, 2013.
[8] A.Sarlan, C.Nadam, S.Basri,“Twitter Sentiment Analysis“, Computer Information Science PETRONAS University of Technology, Perak, Malaysia.
[9] T.Yigitcanlar, N.Kankanamge1 , A.Preston, P.Singh Gill, M.Rezayee, M.Ostadnia, B.Xia and G.Ioppolo, ”How can social media analytics assist authorities in pandemic-related policy decisions? Insights from Australian states and territories”. Health Inf Sci Syst 8, 37 (2020). https://doi.org/10.1007/s13755-020-00121-9.
[10] K. H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin, “Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks”, Kurdistan Journal of Applied Research, vol. 5, no. 3, pp. 54-65, May 2020.
[11] T. Vijay, A. Chawla, B. Dhanka and P. Karmakar, ”Sentiment Analysis on COVID-19 Twitter Data,” 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2020, pp. 1-7, doi: 10.1109/ICRAIE51050.2020.9358301.
[12] R. Lamsal, Design and analysis of a large-scale COVID-19 tweets dataset. Appl Intell (2020). https://doi.org/10.1007/s10489-020-02029-z.
[13] Ahmed W, Vidal-Alaball J, Downing J, Seguı F. L (2020)
[14] COVID-19 situation update worldwide, as of 5 June 2020 [archive], sur European Centre for Disease Prevention and Control.
[15] TextBlob: Simplified Text Processing, https://textblob.readthedocs.io/.
[16] Coronavirus disease (COVID-19) World Health Organization, Situation Report – 132 Data as received by WHO from national authorities by 10:00 CEST, 31 May 2020.
[17] Oracle,WHAT IS A DATA WAREHOUSE?, www.oracle.com/ca-fr/database/what-is-a-data-warehouse .
[18] Analytics and Business Intelligence, www.gartner.com.
[19] Jérôme Darmont, Patrick Marcel. Data warehouses and OLAP, analysis and decision in the company. CNRS Editions. Big Data Uncovered, pp. 132-133, 2017, 978-2-271-11464-8. ffhal-01493948f.
[20] Ronan Tournier, Online document analysis (OLAP), PhD thesis in computer science, under the direction of Gilles Zurfluh, Toulouse, University of Toulouse, 2007.
[21] Maha Ben Kraiem, Jamel Feki, Kais Khrouf, Franck Ravat, Olivier Teste. OLAP4Tweets : Multidimen- sional Modeling of tweets. 19th East-European Conference on Advances in Databases and Information Systems (ADBIS 2015), Sep 2015, Poitiers, France. pp. 68-75. ⟨hal-01343054⟩
[22] X. Liu et al., ”SocialCube : A Text Cube Framework for Analyzing Social Media Data,” 2012 Inter- national Conference on Social Informatics, Lausanne, 2012, pp. 252-259, doi : 10.1109/SocialInforma- tics.2012.87. emergence and transmission. Pathogens. 2016;5(4):66.