Collaborative Business Intelligence Virtual Assistant

Olga Cherednichenko [1] and Fahad Muhammad [1]

[1] *Univ Lyon, Univ_Lyon 2, UR ERIC – 5 avenue Mendès France, 69676 Bron Cedex, France*

Abstract

Current business environment requires new methods that incorporate more intelligent technologies and tools capable to provide fast, accurate and reliable information for decision making. This paper deals with data mining applications. It describes the unified business intelligence semantic model, coupled with a data warehouse and collaborative unit to employ data mining technology. The virtual assistant for collaborative business intelligence is suggested.

Keywords 1

Artificial Intelligence, Collaborative Business Intelligence, Virtual Assistance, Machine Learning

1. Introduction

The decision-making process is complex and, as a rule, depends significantly on the information that the person who makes the decision owns. Today's world is characterized by huge volumes of accumulated data in various domains. These data can be really helpful in terms of preparing and making the decisions. However, this data are collected and stored by different unrelated software systems, stored in different formats, providing different levels of access and security. In addition, these data may be incomplete, contradictory, unreliable. To solve problems associated with the processing of large volumes of data, they turn to data analysts. Business intelligence (BI) helps you gain valuable insights and make strategic decisions. Business intelligence tools analyze historical and current data and present the results in intuitive visual formats. A significant obstacle to achieving the effect of using the accumulated data is the lack of direct communication between technical specialists and decision makers and business process analysts. The solution to this problem is to apply the approach of collaborative business intelligence (CBI).

As the analysis shows, the need for data research arises not only from business in order to increase its profits, or the government to solve national problems, but also from society and individual citizens to understand and justify socially significant or private decision-making. In this case, it is quite difficult to organize the interaction of potential users, decision makers, and technical specialists in data analysis. The project BI4people [1] is created to help solve these problems. The aim of BI4people is to bring the power of Business Intelligence to the largest possible audience, by implementing the data warehousing process in software-as-a-service mode, from multisource, heterogeneous data integration to intuitive analysis and data visualization [2]. The main idea is to collect people in one virtual and encourage them leave their comment or opinions for general purpose. Moreover, reusing another collaborators' results or comments makes general BI - Collaborative BI.

MoMLeT+DS 2023: 5th International Workshop on Modern Machine Learning Technologies and Data Science, June, 3-4, 2023, Lviv-Leiden, Ukraine

EMAIL: olga.cherednichenko@univ-lyon2.fr (O. Cherednichenko); fahad.muhammad@eric.univ-lyon2.fr (F. Muhammad)
ORCID: 0000-0002-9391-5220 (O. Cherednichenko); 0000-0002-7258-9884 (F. Muhammad)

2. Background

Data exploration is an important component of the BI process, which involves collecting, identifying, and analyzing data to discover meaningful insights and patterns. The main goal of data exploration is to identify key business opportunities and challenges that can drive decision-making and improve business performance.

The main idea of our research is to model CBI processes in distributed virtual teams via interaction of user and CBI Virtual Assistant (Fig. 1).

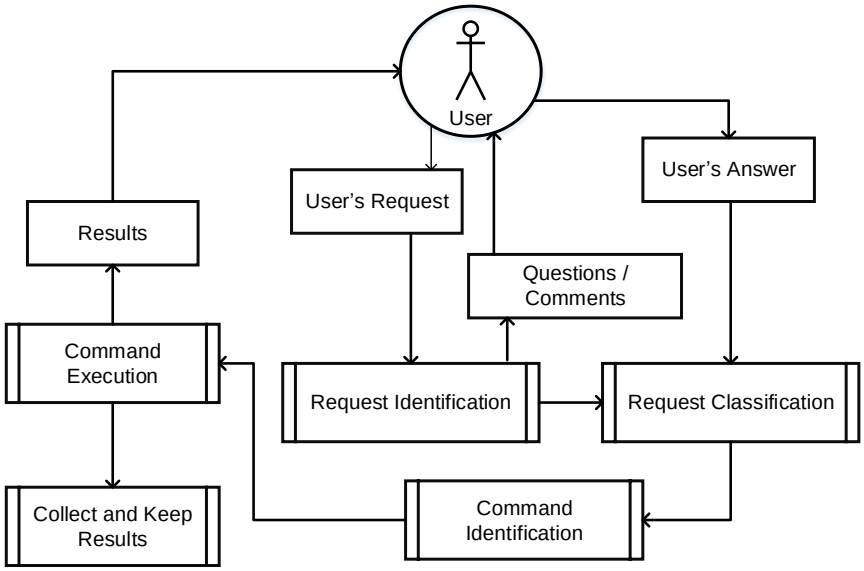

Figure 1: The general workflow

A chatbot is a computer program that mimics human conversation through text or voice interactions with users [3]. Chatbots can provide several advantages for e-commerce businesses, including:

1. 24/7 availability: Chatbots can operate round the clock, providing customers with access to support and information outside regular business hours. This ensures that customers can get their questions answered and issues resolved even if they reach out at odd hours.
2. Faster response times: Chatbots can instantly provide answers to common queries, freeing up support staff to tackle more complex issues. This can result in faster response times and reduced wait times for customers, which can improve their overall experience with the brand.
3. Personalization: Chatbots can use data about the customer's previous purchases, browsing history, and preferences to offer personalized product recommendations and promotions. This can help businesses to build a stronger connection with customers and increase sales.
4. Scalability: Chatbots can handle multiple conversations simultaneously, allowing businesses to handle a large volume of customer queries and support requests without hiring additional staff. This can help businesses to scale their customer support operations more efficiently.
5. Cost-effectiveness: Chatbots can provide cost-effective support and reduce the need for businesses to hire additional support staff. This can help businesses to save money while providing excellent customer service.

The state of the art of CBI can be described as a growing field with an increasing number of tools and techniques being developed to enable effective collaboration among teams in decision-making processes. Some of the most significant advancements in CBI include the integration of social media features, mobile accessibility, and cloud-based solutions. These developments have enabled users to work collaboratively and access data from any location, on any device, at any time. Additionally, the use of NLP and ML has made it easier for users

to interact with data and extract insights, making decision-making processes more efficient and effective. Natural Language Querying (NLQ) can also make it easier for non-technical users to access and analyze data.

The goal of a virtual assistant is to make data exploration more accessible to a wider range of users and to reduce the time and effort required for data analysis. It is an idea of creating innovative CAs is to transform the way users interact with data and ML models and to make data science more accessible to a wider range of users.

3. The state of the art

Chatbots can be classified into different categories based on their functionalities and the type of collaboration they facilitate [4, 5, 6]. They can provide insights, recommendations, or predictions based on the available data. Chatbots can also send notifications and alerts to users triggered by predefined actions, such as changes in data or anomalies in key metrics.

Although a chatbot is a type of conversational agent (CA), not all CAs are chatbots. CA is a broader term that includes any computer program or system that can engage in natural language interactions with users [7, 8]. CAs can be rule-based or use machine learning (ML) and natural language processing (NLP) techniques to comprehend and respond to user inputs.

Drawing upon a review of 233,085 papers, the authors of [9] observed that despite the widespread interest in chatbot integration, only 81 papers met the evaluation criteria for inclusion, such as a relevant abstract, clear methodology presentation, full-text availability, relevance, and use of English. The findings from [9] indicates that "chatbot" and "artificial intelligence" are the two keywords with the highest co-occurrence in the selected papers. The use of the Python programming language is prevalent in developing chatbots [9]. Consequently, we can conclude that while the topic is not novel, it is still cutting-edge, with numerous successful chatbot and conversational agent implementations demonstrating their potential. Various tools and language models are available to implement the personal shopping assistant.

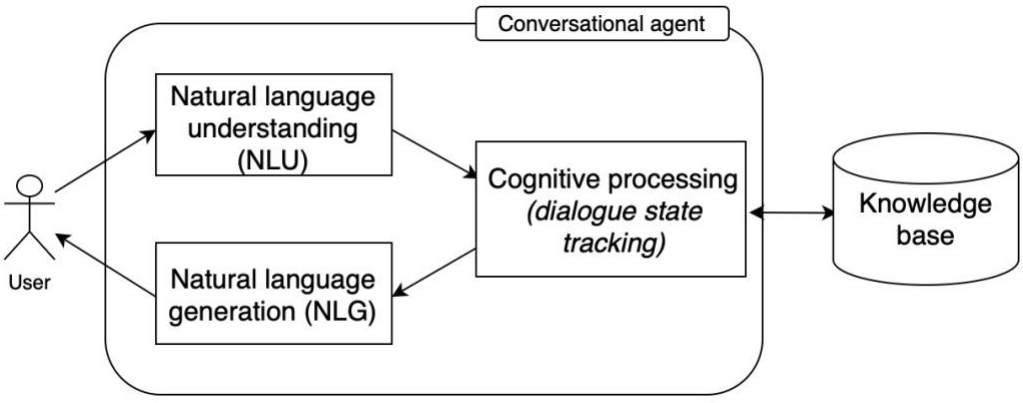

Figure 2: The general pipeline for conversational agent (adopted from [10])

4. Methods and Materials

We propose to consider the following main stages of the research. First, a domain must be defined in which collaborative analysis and BI can be modeled. Considering different goals, preferences, experience and conditions, different users will access the same data with different

requests forming the content of collaborative session. Second, data sources need to be identified. Combining data from different sources requires solving the problems of data consolidation, cleaning, and standardization. Thirdly, one of the main stages is the formation of a knowledge base of collaborative decision-making cases. At this stage, you need to develop an information model for collecting data about each session, including user behavior and the results of his research, as well as interaction with other users. Fourth, it is necessary to develop a convenient interface for visualizing data and organizing interaction in the virtual space. The final stage is associated with the processing, analysis, and summarizing of the collected data about user behavior. We believe that as a result we will be able to create a CBI framework and prove models and technologies for supporting virtual space, which will expand the functionality of the BI4people project platform.

Let us describe the data we use for experimenting. For each bodily accident occurring on a road open to public traffic, involving at least one vehicle and causing at least one victim requiring treatment, information describing the accident is entered by the law enforcement unit (police, gendarmerie, etc.) which intervened at the scene of the accident [11]. These entries are compiled in a form entitled bodily accident analysis report. All of these files constitute the national file of traffic injury accidents known as the "BAAC file" administered by the National Interministerial Road Safety Observatory "ONISR".

The databases, extracted from the BAAC file, list all the bodily injury accidents occurring during a specific year in mainland France and in the overseas departments with a simplified description. This includes accident location information, as entered, as well as information regarding the characteristics of the accident and its location, the vehicles involved and their victims. Every year, road accidents cause thousands of deaths. People are wondering what the causes are, what specific issues influence the most, who are under the risk etc. We use the dataset available at [11] as an example how people can explore data collaboratively and show how the Virtual Assistant can support them. The data consists of four datasets which are describe features of accidents (tabl. 1), places (tabl. 2), users (tabl. 3), and vehicles (tabl. 4).

Table 1

Data set “caracteristics.csv” [11]

Feature	Description	Type	Possible values
Num_Acc	Accident ID	Int	not specified
an	Year of the accident	Int	not specified
mois	Month of the accident	Int	not specified
jour	Day of the accident	Int	not specified
hrmn	Time of the accident in hour and minutes (hhmm)	Int	not specified
lum	Lighting : lighting conditions in which the accident	Int	1 - Full day 2 - Twilight or dawn 3 - Night without public lighting 4 - Night with public lighting not lit 5 - Night with public lighting on
agg	Localisation	Int	1 - Out of agglomeration 2 - In built-up areas
int	Type of Intersection	Int	1 - Out of intersection 2 - Intersection in X 3 - Intersection in T 4 - Intersection in Y 5 - Intersection with more than 4 branches

			6 - Giratory 7 - Place 8 - Level crossing 9 - Other intersection
atm	Atmospheric conditions		1 - Normal 2 - Light rain 3 - Heavy rain 4 - Snow - hail 5 - Fog - smoke 6 - Strong wind - storm 7 - Dazzling weather 8 - Cloudy weather 9 - Other
col	Type of collision	Int	1 - Two vehicles - frontal 2 - Two vehicles - from the rear 3 - Two vehicles - by the side 4 - Three vehicles and more - in chain 5 - Three or more vehicles - multiple collisions 6 - Other collision 7 - Without collision
com	Municipality	Int	The commune number is a code given by INSEE. The code has 3 numbers set to the right
adr	Postal address	Str	variable filled in for accidents occurring in built-up areas
gps	GPS coding	Str	1 original character: M = Métropole A = Antilles (Martinique or Guadeloupe) G = Guyane R = Réunion Y = Mayotte
lat	Latitude	Int	not specified
long	Longitude	Int	not specified
dep	Department	Int	INSEE Code (National Institute of Statistics and Economic Studies) of the department followed by a 0 (201 Corse-du-Sud - 202 Haute-Corse)

Table 2
Data set “places.csv” [11]

Feature	Description	Type	Possible values
Num_Acc	Identifier of the accident	Int	not specified
catr	Road category	Int	1 – Motorway 2 – National road 3 – Departmental road 4 – Communal roads 5 – Outside the public network 6 – Car park open to public traffic 7 – Urban metropolis roads

			9 – other way
voie	Route number	Int	not specified
V1	Numerical index of the road number	Int	not specified
V2	Alphanumeric index letter of the route	Str	not specified
circ	Traffic regime	Int	-1 – Not specified 1 – One way 2 – Bidirectional 3 – With separate carriageways 4 – With variable assignment channels
nbv	Total number of traffic lanes	Int	not specified
pr	Attachment PR number (upstream terminal number)	Int	The value -1 means that the PR is not filled in
pr1	Distance in meters to the PR (compared to the upstream terminal)	Int	The value -1 means that the PR is not filled in
vosp	Indicates the existence of a reserved lane, regardless of whether or not the accident took place on this lane.	Int	-1 – Not filled in 0 – Not applicable 1 – Cycle path 2 – Cycle lane 3 – Reserved lane
prof	Long profile describes the gradient of the road at the location of the accident	Int	- 1 – Not specified 1 – Flat 2 – Slope 3 – Top of hill 4 – Bottom of hill
plan	Plan layout	Int	-1 – Not filled in 1 – Straight part 2 – Curved left 3 – Curved right 4 – “S” shaped
lartpc	Width of the central reservation (TPC) if it exists (in m)	Int	not specified
larrout	Width of the carriageway allocated to the circulation of vehicles does not include hard shoulders, TPCs and parking spaces (in m).	Int	not specified
surf	Surface condition	Int	-1 – Not specified 1 – Normal 2 – Wet 3 – Puddles 4 – Flooded 5 – Snowy 6 – Mud 7 – Icy 8 – Fats – oil

			9 – Other
infra	Planning - Infrastructure	Int	-1 – Not filled in 0 – None 1 – Underground - tunnel 2 – Bridge - flyover 3 – Interchange or connecting ramp 4 – Railway 5 – Developed crossroads 6 – Pedestrian zone 7 – Toll area 8 – Construction site 9 – Others
situ	Situation of the accident	Int	-1 – Not specified 0 – None 1 – On the road 2 – On hard shoulder 3 – On shoulder 4 – On sidewalk 5 – On a cycle path 6 – On another special lane 8 – Other
env1	Maximum authorized speed at the place and at the time of the accident	Int	not specified

Table 3
Data set “users.csv” [11]

Feature	Description	Type	Possible values
Num_Acc	Accident ID	Int	not specified
place	Location of user	Int	the seat occupied in the vehicle by the user at the time of the accident (detail is given by the illustration) 10 – Pedestrian (not applicable)
catu	User category	Int	1 – Driver 2 – Passenger 3 – Pedestrian
grav		Int	
sexe	User gender:	Int	1 – Male 2 – Feminine
trajet	Reason for travel at the time of the accident	Int	-1 – Not specified 0 – Not filled in 1 – Home – work 2 – Home – school 3 – Shopping – purchases 4 – Professional use 5 – Walk – leisure 9 – Other
secu	Character intelligence indicates the presence and use of safety equipment	Int	-1 – Not filled in 0 – No equipment 1 – Belt 2 – Helmet

			3 – Children device 4 – Reflective vest 5 – Airbag (2WD/3WD) 6 – Gloves (2WD/3WD) 7 – Gloves + Airbag (2WD/3WD) 8 – Not determinable 9 – Other
locp	Location of the pedestrian	Int	-1 – Not filled in 0 – Not applicable On pavement: 1 – A + 50 m from the pedestrian crossing 2 – A – 50 m from the pedestrian crossing On pedestrian crossing: 3 – Without light signaling 4 – With light signaling Various : 5 – On sidewalk 6 – On shoulder 7 – On refuge or BAU 8 – On counter aisle 9 – Unknown
actp	Action of the pedestrian	Int	-1 – Not filled in Moving 0 – Not filled in or not applicable 1 – Direction of vehicle hitting 2 – Reverse direction of the vehicle Miscellaneous 3 – Crossing 4 – Hidden 5 – Playing – running 6 – With animal 9 – Other
etatp	This variable makes it possible to specify whether the injured pedestrian was alone	Int	-1 – Not filled in 1 – Alone 2 – Accompanied 3 – In a group
an_nais	User's year of birth	Int	not specified
num_veh	Identification of the vehicle	Str	for each user occupying this vehicle (including pedestrians who are attached to vehicles that hit them) - alphanumeric code

Table 4
Data set “vehicles.csv” [11]

Feature	Description	Type	Possible values
Num_Acc	Accident ID	Int	not specified
senc	Flow direction	Int	-1 – Not filled in 0 – Unknown

			1 – PK or PR or ascending mailing address number 2 – PK or PR or descending postal address number 3 – No mark
catv	Category of vehicle	Int	00 – Indeterminable 01 – Bicycle 02 – Moped <50cm3 03 – Cart (Quadricycle with bodied motor) (formerly "cart or motor tricycle") 04 – Reference unused since 2006 (registered scooter) 05 – Unused reference since 2006 (motorcycle) 06 – Reference unused since 2006 (sidecar) 07 – LV only 08 – Reference unused since 2006 (VL + caravan) 09 – Reference unused since 2006 (VL + trailer) 10 – LCV only 1.5T <= GVW <= 3.5T with or without trailer (formerly LCV only 1.5T <= GVW<= 3.5T) 11 – Reference unused since 2006 (VU (10) + caravan) 12 – Reference unused since 2006 (VU (10) + trailer) 13 – PL only 3.5T <PTCA <= 7.5T 14 – PL only > 7.5T 15 – HGV > 3.5T + trailer 16 – Road tractor alone 17 – Road tractor + semi-trailer 18 – Reference unused since 2006 (public transport) 19 – Reference unused since 2006 (tramway) 20 – Special gear 21 – Agricultural tractor 30 – Scooter < 50 cc 31 – Motorcycle > 50 cm3 and <= 125 cm3 32 – Scooter > 50 cm3 and <= 125 cm3 33 – Motorcycle > 125 cm3 34 – Scooter > 125 cc 35 – Light quad <= 50 cc (Unbodied motor quadricycle) 36 – Heavy quad > 50 cm3 (Quadricycle with motor without bodywork) 37 – Buses

			38 – Bus 39 – Train 40 – Tramway 41 – 3WD <= 50cc 42 – 3WD > 50cc <= 125cc 43 – 3WD > 125 cc 50 – Motor EDP 60 – EDP without engine 80 – VAE 99 – Other vehicle
occutc	Number of occupants in public transport	Int	not specified
obs	Fixed obstacle struck	Int	-1 – Not specified 0 – Not applicable 1 – Parked vehicle 2 – Tree 3 – Metal slider 4 – Concrete slide 5 – Other slide 6 – Building, wall, bridge pier 7 – Vertical signaling support or emergency call station 8 – Pos 9 – Street furniture 10 – Parapet 11 – Island, refuge, high boundary 12 – Sidewalk curb 13 – Ditch, embankment, rock face 14 – Other fixed obstacle on roadway 15 – Other fixed obstacle on sidewalk or shoulder 16 – Obstacle-free road exit 17 – Nozzle – aqueduct head
obsm	Moving obstacle struck	Int	-1 – Not specified 0 – None 1 – Pedestrian 2 – Vehicle 4 – Rail vehicle 5 – Domestic animal 6 – Wild animal 9 – Other
choc	Initial shock point	Int	-1 – Not specified 0 – None 1 – Before 2 – Front right 3 – Front left 4 – Back 5 – Right back 6 – Left Rear 7 – Right side 8 – Left side 9 – Multiple shocks (barrels)

manv	Main maneuver before the accident	Int	-1 – Not specified 0 – Unknown 1 – Without change of direction 2 – Same direction, same lane 3 – Between 2 lines 4 – In reverse 5 – Against the grain 6 – By crossing the central reservation 7 – In the bus lane, in the same direction 8 – In the bus lane, in the opposite direction 9 – By fitting in 10 – By making a U-turn on the roadway changing lanes 11 – Left 12 – Right Deported 13 – Left 14 – Right Turning 15 – Left 16 – Right Exceeding 17 – Left 18 – Right Various 19 – Crossing the roadway 20 – Parking maneuver 21 – Avoidance maneuver 22 – Door opening 23 – Stopped (excluding parking) 24 – Parked (with occupants 25 – Driving on sidewalk 26 – Other maneuvers engine
num_veh	Identification of the vehicle	Str	for each user occupying this vehicle (including pedestrians who are attached to vehicles that hit them) - alphanumeric code

Standard metadata formats exist to facilitate their collection, search and automatic processing. The retained metadata are as follows:

- Title
- Acronym
- Description
- Licence
- Update frequency
- Key words
- Temporal coverage
- Spatial coverage
- Spatial granularity
- Private mode

On the one hand, data reusers struggle to identify quality datasets and to assess whether such and such a dataset is worthy of interest. On the other hand, data producers are not

sufficiently encouraged and supported to improve the quality of their data. It is set up a metadata quality score on data.gouv.fr.

The table 5 is depicts the metadata used.

Table 5
Table title

Criteria	Description
Description of data	The description of the data is of high quality (the description of the data set is sufficiently long).
Update	- The update frequency is entered. - The update frequency is respected
License	- The license is populated. - The license is open
Resource Metadata	Presence of at least one resource with a declared open format
Spatial coverage	- Spatial coverage is provided - The spatial granularity is filled in
Temporal coverage	The temporal coverage of the data is entered

5. Results

In 2015, when the town hall of Paris launched its Cycling Plan, it could not have imagined that the end of it would coincide with a health context favoring its utilitarian practice. If the first plan was ambitious in its redevelopment of the city's cycle paths, the balance sheet of bicycle accidents in Paris and the reasons relating to it question the effectiveness of the first plan.

The advent of the health crisis in 2020 relating to the Covid-19 pandemic favors cycling but requires these recent developments to be maintained and expanded. The appearance of the so-called "coronapists" with the aim of improving traffic flow and relieving public transport has initiated many new cyclists. If many European cities like Saint-Etienne or Marseille decide to erase them after a few weeks, the town hall of Paris obtains the agreement of the government that they are supported in the plan “France Relance”. 2020 is becoming the year of the bike. The culture of utilitarian cycling is anchored in the daily lives of many Parisians, questioning their safety.

The proportion of accident victims wearing a helmet is also the majority and raises the question of prevention and risky behavior adopted by cyclists in Paris. Is it due to an infrastructure problem that supports the idea that the roads are not safe enough, even for users aware of the risks? How effective are city hall's prevention efforts? The data does not allow us to determine the causes of the accidentology, nevertheless they shed light on persistent problems.

Based on the analysis done, there are several important features that must be implemented in virtual assistant software in order to assist novice buyers effectively. These features include:

- The ability to perform various data exploration commands such as filtering, querying, selecting, and setting parameters.
- An information retrieval module is necessary to find relevant information, explore options, and research user needs.
- Item matching is necessary to compare different offers and proposals.

- Personalization based on user preferences, history, and behavior, which can enhance the relevance and effectiveness of the recommendations provided by the assistant.
- Machine learning algorithms to continuously learn from user interactions and improve the assistant's ability to provide personalized recommendations.
- Language understanding and text generation are both necessary in order to effectively communicate with the user.

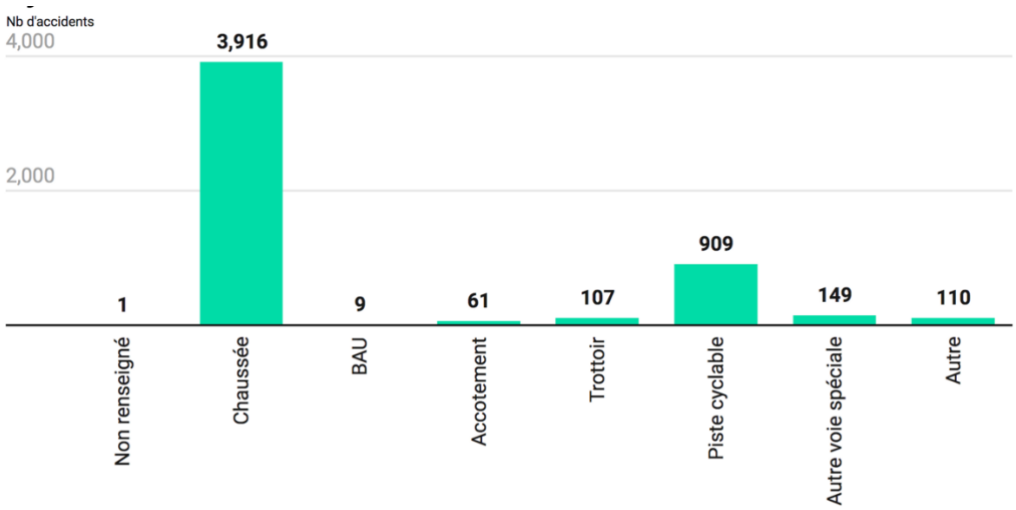

Figure 3: Roads much more dangerous than tracks cycle paths

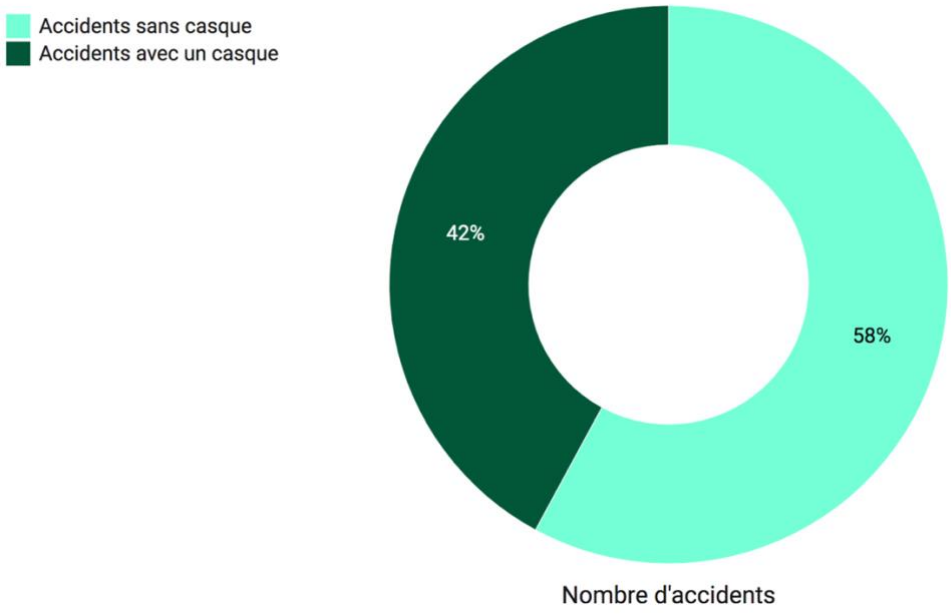

Figure 4: Roads much more dangerous than tracks cycle paths

6. Discussion and Conclusion

In this paper we have presented an approach how to create, approbate and estimate collaborative decision-making models. BI systems are vastly used as a tool to support decision-making in different kind of organizations. CBI give even more opportunities for reasonable

decision-making as they allow using external information from various sources. We are collecting and processing data, developing convenient interface and tools for collaborative analysis. The next step is to implement a prototype

7. Acknowledgements

The research study depicted in this paper is funded by the French National Research Agency (ANR), project ANR-19-CE23-0005 BI4people (Business intelligence for the people).

8. References

[1] Business intelligence for the people. URL: https://eric.univ-lyon2.fr/bi4people/index-en.html.

[2] Muhammad F., Darmont, J., Favre C. (2022) The Collaborative Business Intelligence Ontology (CBIOnt), 18e journées Business Intelligence et Big Data (EDA -22), Clermont-Ferrand, Octobre 2022; RNTI, Vol. B-18

[3] D. Power, R. Sharda, Business Intelligence and Analytics, Wiley, 2015. doi:10.1002/9781118785317.weom070011.

[4] J. Masche, N.-T. Le, A review of technologies for conversational systems, in: International Conference on Computer Science, Applied Mathematics and Applications, Springer, 2017, pp. 212–225.

[5] N. Svenningsson, M. Faraon, Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots, in: Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference (AICCC 2019), Association for Computing Machinery, New York, NY, USA, 2020, pp. 151–161. doi:10.1145/3375959.3375973.

[6] G. Caldarini, S. Jaf, K. McGarry, A Literature Survey of Recent Advances in Chatbots, Information 13(1) (2022) 41. doi:10.3390/info13010041.

[7] M. Akhtar, J. Neidhardt, H. Werthner, The Potential of Chatbots: Analysis of Chatbot Conversations, in: 2019 IEEE 21st Conference on Business Informatics (CBI), IEEE, 2019, pp. 397–404. doi:10.1109/CBI.2019.00052.

[8] R. Bavaresco, D. Silveira, E. Reis, J. Barbosa, R. Righi, C. Costa, R. Antunes, M. Gomes, C. Gatti, M. Vanzin, S. C. Junior, E. Silva, C. Moreira, Conversational agents in business: A systematic literature review and future research directions, Computer Science Review 36 (2020) 100239.

[9] Gamboa-Cruzado, J., et al. Use of chatbots in e-commerce: a comprehensive systematic review. Journal of Theoretical and Applied Information Technology. 101.4 (2023). http://www.jatit.org/volumes/Vol101No4/3Vol101No4.pdf

[10] Wahde, Mattias & Virgolin, Marco. (2022). Conversational Agents: Theory and Applications. https://arxiv.org/pdf/2202.03164.pdf

[11] Bases de données annuelles des accidents corporels de la circulation routière - Années de 2005 à 2021. https://www.data.gouv.fr/en/datasets/bases-de-donnees-annuelles-des-accidents-corporels-de-la-circulation-routiere-annees-de-2005-a-2021/#_

[12]