

Text Mining Approaches Oriented on Customer Care Efficiency

Alessandro Massaro ¹, Nicola Magaletti ¹, Gabriele Cosoli ¹, Vito Giardinelli ¹, and
ANGELO LEOGRANDE ²

¹Affiliation not available

²LUM ENTERPRISE S.R.L.; LUM UNIVERSITY GIUSEPPE DEGENNARO

October 30, 2023

Abstract

In the proposed work is performed a text classification for a chatbot application used by a company working in assistance services of automatic warehouses. industries. Specifically, text mining technique is adopted for the classification of questions and answers. Business Process Modeling Notation (BPMN) models describe the passage from “*AS-IS*” to “*TO BE*” in the context of the analyzed industry, by focusing the attention mainly on customer and technical support services where chatbot is adopted. A two-step process model is used to connect technological improvements and relationship marketing in chatbot assistance: the first step provides the hierarchical clustering able to classify questions and answers through Latent Dirichlet Allocation -LDA- algorithm, and the second one executes the Tag Cloud representing the visual representation of more frequent words contained in the experimental dataset. Tag cloud is used to show the critical issues that customers find in the usage of the proposed service. By considering an initial dataset, 24 hierarchical clusters are found representing the preliminary combination of the couple’s question-answer. The proposed approach is suitable to automatically construct a combination of chatbot questions and appropriate answers in intelligent systems.

Text Mining Approaches Oriented on Customer Care

Efficiency

Abstract

In the proposed work is performed a text classification for a chatbot application used by a company working in assistance services of automatic warehouses. industries. Specifically, text mining technique is adopted for the classification of questions and answers. Business Process Modeling Notation (BPMN) models describe the passage from “*AS-IS*” to “*TO BE*” in the context of the analyzed industry, by focusing the attention mainly on customer and technical support services where chatbot is adopted. A two-step process model is used to connect technological improvements and relationship marketing in chatbot assistance: the first step provides the hierarchical clustering able to classify questions and answers through Latent Dirichlet Allocation -LDA- algorithm, and the second one executes the Tag Cloud representing the visual representation of more frequent words contained in the experimental dataset. Tag cloud is used to show the critical issues that customers find in the usage of the proposed service. By considering an initial dataset, 24 hierarchical clusters are found representing the preliminary combination of the couple’s question-answer. The proposed approach is suitable to automatically construct a combination of chatbot questions and appropriate answers in intelligent systems.

Keywords: *Chatbot, Speech Recognition, Natural Language Processing-NLP, Hierarchical Clustering, Business Process Management and Notation-BPMN.*

1. Literature Review

Authors in (Sari, et al., 2020) consider the use of chatbots to promote business development creating the conditions to improve the customer relationships. The efficacy of chatbots in Business to Business -B2B- is oriented to increase the level of Customer Experience -CX-, which is strictly connected to the ability of customers to use the technology (Kushwaha, et al., 2021) . Chatbots that can apply self-learning are also able to promote a higher degree of CX. If customers increase their trust about chatbot in business services, then it will be diffused in a widespread way (Przegalinska, et al., 2019). The use of chatbot in combination with Customer Relationship Management -CRM- can improve the ability of the company to offer information towards its clients (Setiawan, et al., 2019). The efficacy

of chatbots is a function of the Natural Language Processing -NLP- algorithm adopted for the development (Chao, et al., 2021): Convolutional Neural Networks-CNN- and Long Short-Term Memory-LSTM- (Massaro, et al., 2021) algorithms can be applied to optimize the recognition of words and phrases and a combination of questions and answers. The application of NLP in chatbot systems can also improve the efficacy of health assistance as for telemedicine applications (Omeregbe, et al., 2020). Chatbots are important also in strategic marketing enforced by sentiment analysis and customer experience (Sidaoui, et al., 2020). Authors in (Dodero, et al., 2020) propose a methodology based on NLP, to augment the ability of chatbots in understanding the context in which the conversation is realized. Chatbot can be adopted as a semi-supervised modality to contextualize responses in the minimization of prior data, by improving customer satisfaction (Paul, et al., 2019). Other software tools can evaluate the chatbot-user dialogue under a qualitative point thus obtaining statistics to be used to augment the efficacy of the customer service (Følstad, et al., 2021). Chatbot is able not only to learn from customers' responses but also to generate inferences about the customers' behaviors, intentions and needs through the dynamic application of knowledge in customer care (Pantano & Pizzi, 2020). Chatbot can also operate as a virtual front-office (Massaro, et al., 2018), and can be optimized by sentiment analysis (Massaro, et al., 2019), (Al Islami, et al., 2020). It is also possible to implement a chatbot system able to recognize words directly in speeches using LSTM algorithms (Massaro, et al., 2020). In human resource sector, chatbots can be used as tools for recruitment (Chou, et al., 2019). In *non-business activities*, chatbot is applied in healthcare (Maeda, et al., 2019), (Kowatsch, et al., 2017), and to assist students in their undergraduate and postgraduate courses (Murad, et al., 2019), (Hien, et al., 2018), (Dharaniya, et al., 2020), (Ashok, et al., 2021), (Qaffas, 2019). The usage of chatbot can solve the question of the scarcity of human resources in medical services in respect to patients' demand (Tjiptomongsoguno, et al., 2020). Chatbot can improve the offering of services in the health industry (Sheth, et al., 2019). Chatbots are efficiently used also in public administration (Lommatzsch, 2018) and transportation (Dharani, et al., 2020). Furthermore, Automatic Speaker Recognition-ASR- tools (Prasad & Ranjith, 2020) are integrable for voice-based chatbot systems thus providing a technological advance. Chatbots are suitable to interpret the various typologies of emotions in healthcare (Luo, et al., 2021), by improving the ability of medical systems to process health information (Madhu, et al., 2017), (Mathew, et al., 2019), (Yorita, et al., 2020), (Siangchin & Samanchuen, 2019).

Concerning the topic of the paper of the improvement of customer assistance, in (Nagarhalli, et al., 2020) are shown how conversational systems can be used to develop assistance services in various domains. In addition, artificial intelligence is potentially useful in supporting consumers about shopping (Pizzi, et al., 2021). In the manufacturing sector, chatbots can be used for technical

assistance (Mantravadi, et al., 2020). By considering e-commerce platforms, the online chatbot is a good channel for customer communication and assistance (Landim, et al., 2021). Chatbot is applied in other industrial sectors such as the luxury fashion sector (Chung, et al., 2020). Different advantages in assistance services for customers can be achieved by chatbot (Behera, et al., 2021). Chatbot has a customer strategic role because it can be based on a conversation strategy and recommendations systems to improve customer assistance (Ikemoto, et al., 2018). The state of the art highlights the importance of chatbot in industrial applications more specifically for customer care processes.

The entire project of implementation of an AI-Chatbot has been realized in the context of the Smart District 4.0 project -SD4.0. The SD4.0 project has been founded by the Italian Minister of Economic Development -MISE to promote innovation technology and research and development in Small and Medium Enterprises -SMEs.

The paper is structured as follows: the second paragraph contains the Business Process Model and Notation-BPMN used to optimize the function of the Chatbot system for the specific case study following economic and financial advantages, the third paragraph describes in detail the two-step process model to connect technological improvements and relationship marketing in chatbot assistance, the fourth paragraph presents the results of hierarchical clustering, the fifth paragraph shows the results of the Latent Dirichlet Allocation-LDA algorithm used to generate the tag cloud.

1. Business Process Model and Notation-BPMN and the Economic and Financial Advantages of the AI-Chatbot

In our analysis, the application of the AI-Chatbot consent to realize the following goals that are valuable in the sense of Customer Process Management:

- *Queries pre-classification through the application of hierarchical clustering:* through the application of the hierarchical clustering it is possible to verify the presence of questions that are recurrent among customers; those questions can shed light on some non-performing elements of the product; that information can be used to promote a process of re-invention and re-industrialization of services that can better serve the necessity of the customer.
- *Answer pre-classification through the application of hierarchical clustering:* the analysis of the answers through the application of the hierarchical clustering is necessary to verify if the Chatbot is effectively able to promote the right answers to customers; specifically, the AI-Chatbot is supposed to learn from the dynamic of questions and answers and so, marginally, its answers should change reflecting relevant modifications in the question of the customers.

- *Automated customer care through the application of the KB Virtual Assistant:* the application of the automated customer care can improve the quality of the relationship with the customer; specifically, the AI-Chatbot does not only reply to questions based on its inner logic, but on the contrary is programmed, through NLP, to learn and offer change-oriented solutions even in the case of very similar questions.

These three elements are associated with the Product Quality Management Process i.e. they have a positive impact on the ability of the product to generate value for customers. The presence of an AI-Chatbot that is oriented to promote services in the context of the company-customer relationships, offers a series of economic and financial advantages such as:

- *Customer Retention:* is the ability of the firm to develop long run relationships with customers. In the case of retention, the customer does not change the firm that supplies products or services, does not buy similar products and services from competitors. In this case the application of the AI-Chatbot improves the level of quality of services and creates the condition to develop long run relationships between firms and customers.
- *Customer Care:* is the supply of services to customers during and after the purchase of the goods or services. The application of the AI-Chatbot improves the ability of the firm to generate assistance towards its clients creating the conditions for a more performing ability of customer care.
- *Service Quality Performance > Customer Expectations:* is the definition of quality in the service. The AI-Chatbot based on its ability to learn and understand the questions posed by customers through the application of Natural Language Processing-NLP algorithms can effectively over perform the assistance service in respect to customer expectations.
- *Customer loyalty:* There is a positive impact of the application of AI-Chatbot on customer loyalty since the customer can effectively ask to the Chatbot system, how to solve its problems in the usage of the automatic warehouses.

These positive outcomes show the relevance of the application of the AI-Chatbot in the context of the industrial services, to promote a better relationship with customers and create the conditions to improve the Product Quality Management Process. The AI-Chatbot for its ability to learn from questions and answer can promote dynamic dialogue with customers improving either their ability to use efficiently the automatic warehouse, either their trust in the firm, its products, and its customer assistance.

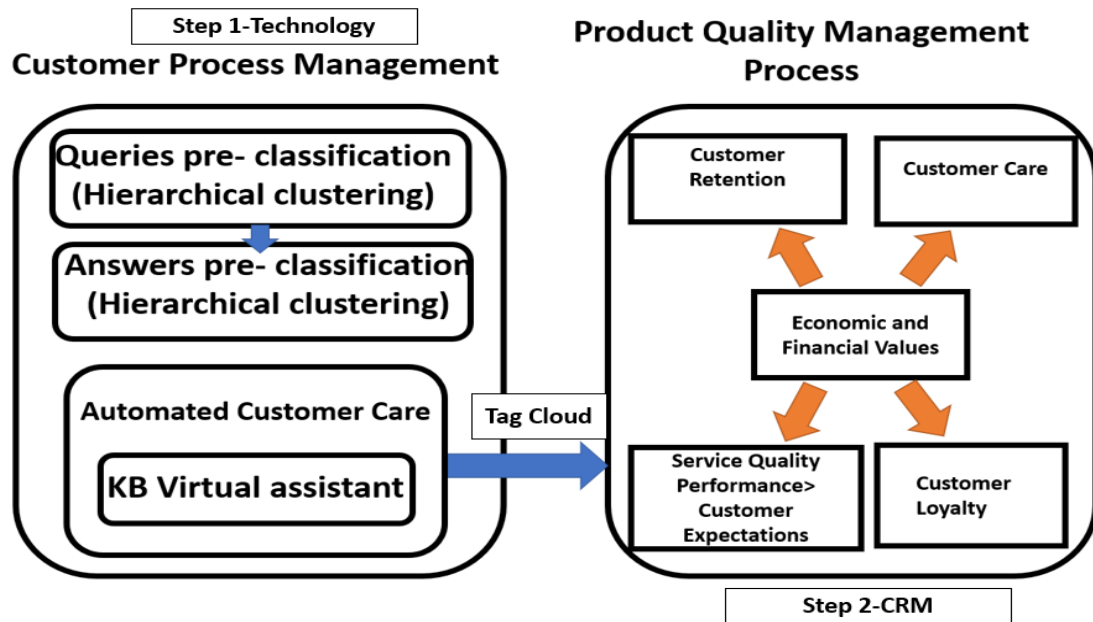


Figure 1. Customer process management and product quality management in the case of AI-Chatbot.

Specifically, the application of the Chatbot in the case of ICAM srl company, either for its impact on the Customer Process Management, either for its ability to improve Product Quality Management, generates an organizational change that is oriented to have significant impacts on the following areas:

- Increase in the number of customers serviced:** the application of the virtual technical assistance improves the ability to serve a greater number of customers with a greater efficiency. The entire customer assistance service is realized using the phone in the AS-IS. The number of customers that a single worker can serve is limited in confrontation with the possibility offered by the application of the AI chatbot. In the “TO BE” version of the service the application of the automated virtual assistance improves the ability to serve a greater number of clients. In this sense ICAM srl can reduce the costs of workers employed in service assistance and increase the number of requests fulfilled. The company can acquire new margins of productivity, improve efficiency, and increase the ability to create a customer relationship management oriented to the promotion of qualitative services.
- Impact on human resources:** in the AS-IS of the firm ICAM the entire process of customer assistance and customer care is offered by human resources that are employed in routinary tasks. The application of the AI-Chatbot in the TO-BE can reduce the human workforce employed in customer care. Workers can be better oriented to realize more complex tasks such as the training of the artificial intelligence and the data analysis oriented to disentangle the customer behavior. In the change from to the “AS-IS” to the “TO-BE” the firm has the

possibility to re-evaluate the human capital and to promote a process of digitalization that can transform a front-end worker into an analyst of customer behavior using data science.

- **Impact on supply chain:** using the AI-Chatbot the firm also can select implicitly the suppliers. In effect when customers complain about the malfunctioning of the product then the firm can recognize recurrent errors in the supply. Using the requests of the clients it is possible to verify the functioning of the automatic warehouse, understanding the more critical issues, and the weaknesses of the system either in the sense of the hardware infrastructure or in the sense of software functionality. The customer experience can offer relevant information to optimize the functioning of the automatic warehouses and can shed lights on some functionality that underperform in respect to the customer expectations. Furthermore, it is possible to use the information relative to the customer experience to evaluate the customer's innovativeness and the ability of clients to use the technological infrastructure used to the management of the automated warehouse.

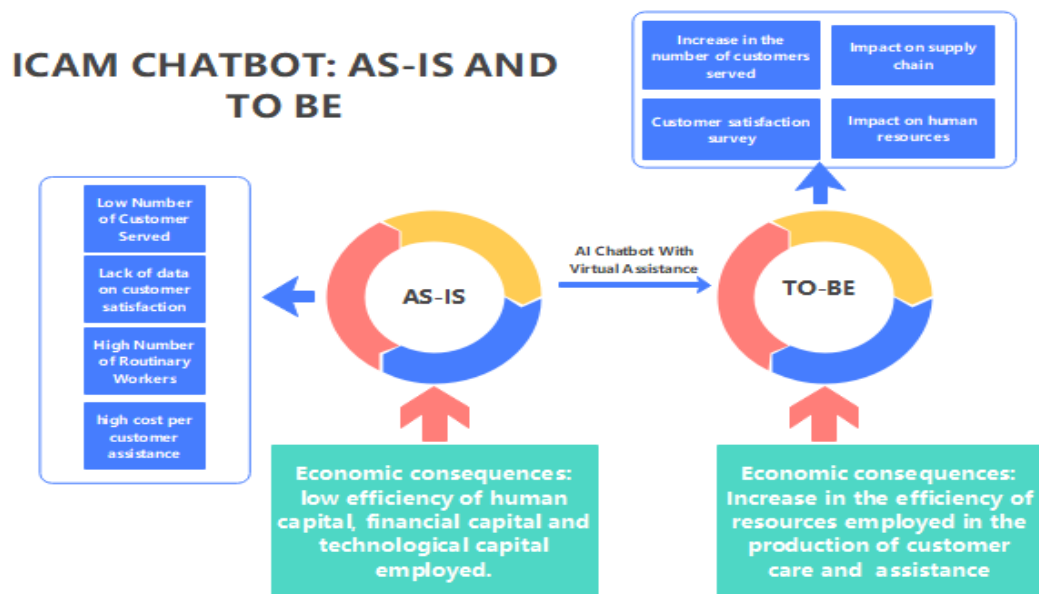


Figure 2. The organizational change between AS-IS and TO-BE in the case of AI-Chatbot for the case study of ICAM.

- **Customer satisfaction survey:** introduce a customer satisfaction survey tool, to investigate any causes of dissatisfaction to eliminate them and improve the technical assistance service, minimizing the number of dissatisfied customers. It might be useful to record all interactions between customers and the company through the platform. The possibility to acquire data that can be used to an effective identification of the clients is relevant to produce either better automatic warehouses or to improve the efficiency of the service assistance. Data can show customer behaviors, customer implicit needs, and can reveal the levels of customer

satisfaction and customer experience. Through the usage of data, the company can promote better products and services and can also verify the degree of efficacy of the targeting of the customers realized during the phase of business planning and programming.

We can use analytical tools to describe the characteristics of organizational change produced in association with the implementation of the ICAM-Chatbot through the application of the Business Process Model and Notation-BPMN (Massaro, 2021), (Massaro & Galiano, 2020).

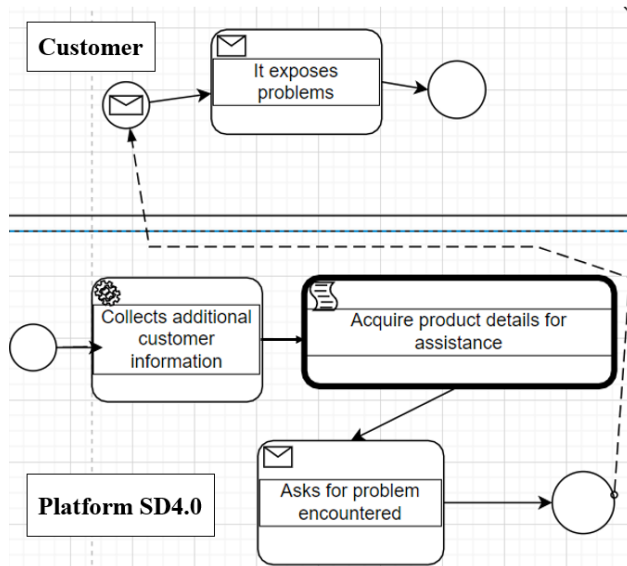


Figure 3. Technical Support Service extracted from the whole "TO BE" processes (see Appendix).

Two sub-processes are presented either for the Sales Customer Service and the Technical Support Service. In the Figure 3 there is a representation of a subprocess of the Technical Support Service. The sub-process starts with the Platform SD4.0 that collects additional customer information. The Platform acquires product details for assistance and asks for problem encountered. Then the platform SD4.0 sends a message to customers that replay explaining the problem. Later the sub-process ends.

In figure 4 there is a representation of a subprocess of Sales Customer Service. The platform SD4.0

verifies the activation of the sub-process. In the following step two alternatives are feasible since there is a macro-process activation:

- *No*: in this case the platform SD4.0 asks for product service information;
- *Yes*: in this case the platform SD4.0 collects additional customer information and later asks for product service information.

After having asked for product service information the analyzed sub-process can follow two different paths:

- The platform SD4.0 can acquire information about the service;
- The platform SD4.0 can send a message to the customer that enters information about the service/product and at this point there are two alternatives: firstly the process ends, secondly

the customer sends a message to platform SD4.0 to acquire information about the product service.

In both cases the usage of the Business Process Management and Notation-BPMN helps the firm to describe and activate the analyzed services. The relationship between the Platform SD4.0 and the Customer is relevant to produce the services that the firm has interested to offer to their clients. Specifically, the passage from the “*AS IS*” to the “*TO BE*” shed lights on the ability of the company to optimize the customer assistance either in the case of “*Sales Customer Service*” either in the case of “*Technical Support Service*”. The whole processes “*AS-IS*” and “*TO-BE*” belonging the extracted processes is described in the appendix. In the proposed “*TO BE*” scenario is applied the two-step methodology described in the second paragraph.

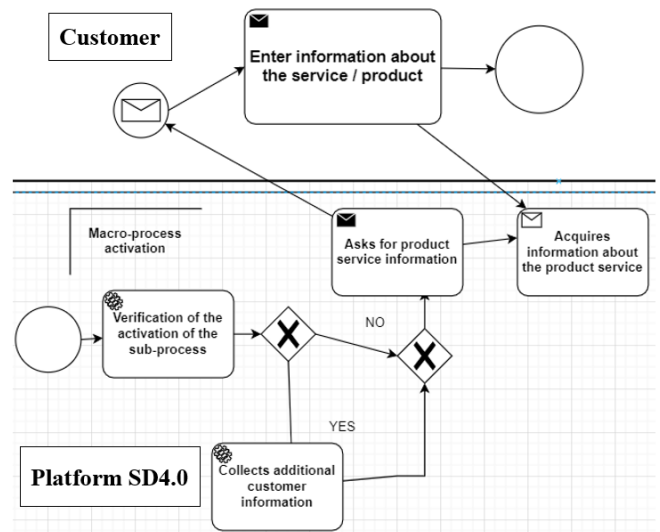


Figure 4. Sales Customer Service-extracted from the whole “*TO BE*” processes (see Appendix).

2. A Two-Step Process Model to Connect Technological Improvements and Relationship Marketing in Chatbot Assistance

A two-step chatbot text mining process model is designed: the first step is based on hierarchical clustering and the second one is addressed on the optimization of relationship marketing through the analysis of Tag Cloud. While in the first step the hierarchical clustering has been used to obtain better technological solutions that are realized through the application of artificial intelligence, machine learning and NLP, in the second step the Tag Cloud has been adopted to identify the most frequent words in the datasets based on the assumption that these expressions reflect the most relevant question for customers. Tag Cloud is able in this sense to offer a graphical representation of the frequency of words in the dataset. Specifically, through the analysis of the Tag Cloud it is possible to verify which are the parts of the automatic warehouse that are more difficult to use for the customer. The analysis of Tag Cloud offers the possibility to verify the ability of the company to promote some relevant goods such as customer service, customer care, customer loyalty, and customer retention.

While in the first step through hierarchical clustering are analyzed questions and answers, in the second step only answers are analyzed. In this way it is possible to verify, through the Tag Cloud, which are the major functionalities that can be implemented to solve customers’ problems, or which are the functionalities that customers find harder to understand.

3. Hierarchical clustering

We perform a hierarchical clustering using Konstanz Miner -KNIME- and an experimental dataset of 65 questions and 65 answers. In the analysis are consider not only the quantitative analysis based on the number of clusters, but also the qualitative analysis relative to the association between clusters and questions-answers. Specifically, the workflow used for the hierarchical cluster analysis is based on four different processes i.e.:

- **Data import:** is composed by tree different KNIME nodes that are “*Excel Reader*”, “*String to Document*” and “*Column Filter*”;
- **Preprocessing:** is based on seven KNIME nodes and one metanode. The different seven KNIME nodes are “*Markup Tag Filter*”, “*Punctuation Erasure*”, “*N Chart Converter*”, “*Stop Word Filter*”, “*Stanford Lemmatizer*”. The metanode is based on the following “*Extract Table Dimension*”, “*Java Edit Variable*”, “*Bag of Words Creator*”, “*Term to String*”, “*Group By*”, “*Row Filter*”, “*Reference Row Filter*”, “*TF*”.
- **Transformation:** that is based on “*Document Vector*” e “*Category Class*”;
- **Clustering:** that is based on the three different clusters that are repeated three times respectively with Euclidean, Manhattan and Cousin methodology that are: “*Hierarchical Clustering*”, “*Hierarchical Cluster View*”, “*Distance Matrix*”, “*Column Filter*”, “*k-Medoids*”, “*Hierarchical Cluster Assigner*”.

KNIME Nodes for Hierarchical Clustering	
Macro process	Nodes
Data Import	Is the first macro-process that is used to introduce data in KMINE. The Data Import is based on the following nodes that are: “ <i>Excel Reader</i> ”, “ <i>String to Document</i> ”, “ <i>Column Filter</i> ”.
Preprocessing	Is a macro process that is based on 7 nodes and one metanode. The seven nodes are indicated as follows: “ <i>Markup Tag Filter</i> ”, “ <i>Punctuation Erasure</i> ”, “ <i>N Chart Converter</i> ”, “ <i>Stop Word Filter</i> ”, “ <i>Stanford Lemmatizer</i> ”. The metanode is realized with the following nodes i.e. “ <i>Extract Table Dimension</i> ”, “ <i>Java Edit Variable</i> ”, “ <i>Bag of Words Creator</i> ”, “ <i>Term to String</i> ”, “ <i>Group By</i> ”, “ <i>Row Filter</i> ”, “ <i>Reference Row Filter</i> ”, “ <i>TF</i> ”.

Transformation	Is based on two different nodes that are “ <i>Document Vector</i> ” and “ <i>Category Class</i> ”.
Clustering	The clusterization process is based on “ <i>Hierarchical Clustering</i> ”, “ <i>Hierarchical Cluster View</i> ”, “ <i>Distance Matrix</i> ”, “ <i>Column Filter</i> ”, “ <i>k-Medoids</i> ”, “ <i>Hierarchical Cluster Assigner</i> ”.

Table 1. KNIME Nodes for Hierarchical Clustering.

Our results show that there are 24 hierarchical clusters that preset the following characteristics:

- *One cluster with 14 singular element;*
- *Six clusters with two elements;*
- *One cluster with three elements;*
- *One cluster with seventeen elements;*
- *One cluster with ten elements;*
- *One cluster with seven elements.*

In Figure 5 we have the representation of the workflow that has been used to run the hierarchical cluster analysis. As we can see in the green part of the workflow that is named “*clustering*” are present three similar series of nodes that in effect are oriented to compute the hierarchical clustering using three different typologies of distances that are:

- *Euclidean Distance;*
- *Manhattan Distance;*
- *Cosine Distance;*

with increasing number of elements offer more information in respect to cluster with lower element to describe the customer requests. This analytic approach can be generalized for any FAQ system.

Row43	Cluster	Question	Answer
Row42	Cluster 1	Come ripristinare l'extracorsa Y basso?	ModalitàManuale
		Come ripristinare l'extracorsa X software?	ModalitàManuale
Row41			
Row40			
Row39			
Row38			
Row37			
Row36			
Row35			
Row34			
Row33			
Row32			
Row31			
Row30			
Row29			
Row28			
Row27			
Row26			
Row25			
Row24			
Row23			
Row22			
Row21			
Row20			
Row19			
Row18			
Row17			
Row16			
Row15			
Row14			
Row13			
Row12			
Row11			
Row10			
Row9			
Row8			
Row7			
Row6			
Row5			
Row4			
Row3			
Row2			
Row1			

Figure 6. The composition of clusters. On the left side with the indication of Row there are the elements of the clusters. For each row there is a correspondent set of questions-answers.

Row43			
Row42			
Row41			
Row40			
Row39			
Row38			
Row37			
Row36			
Row35			
Row34			
Row33			
Row32			
Row31			
Row30			
Row29			
Row28			
Row27			
Row26			
Row25			
Row24			
Row23			
Row22			
Row21			
Row20			
Row19			
Row18			
Row17			
Row16			
Row15			
Row14			
Row13			
Row12			
Row11			
Row10			
Row9			
Row8			
Row7			
Row6			
Row5			
Row4			
Row3			
Row2			
Row1			

Figure 7. The composition of clusters. On the left side with the indication of Row there are the elements of the clusters. For each row there is a correspondent set of questions-answers.

4. Tag Cloud as a Tool to Optimize Relationship Marketing

In addition to the hierarchical clusters, we realize a representation of the dataset with the usage of tag clouds. Specifically, the workflow to obtain the outputs of tag cloud is based on the following workflows:

- *Inputs*: is a macro process based on two different nodes that are “*Excel Reader*” and “*Strings to Document*”.
- *Text Elaboration*: is the main part of the workflow that is based on the following nodes that are “*Punctuation Erasure*”, “*Number Filter*”, “*N Chars Filter*”, “*Stop Word Filter*”, “*Stanford Tagger*”, “*Stanford Lemmatizer*”, “*Case Converter*”, “*Tagged Document Viewer*”;
- *Regulator*: is composed of a single node that is “*Topic Extractor-Parallel LDA*”;
- *Outputs*: is the macro-process that can be used to verify the presence of outputs that is “*Color Manager*” and “*Tag Cloud*”.

The main objective of the realization of a workflow based on tag cloud consists in the final plotting of a word cloud as a graphical representation of the answers.

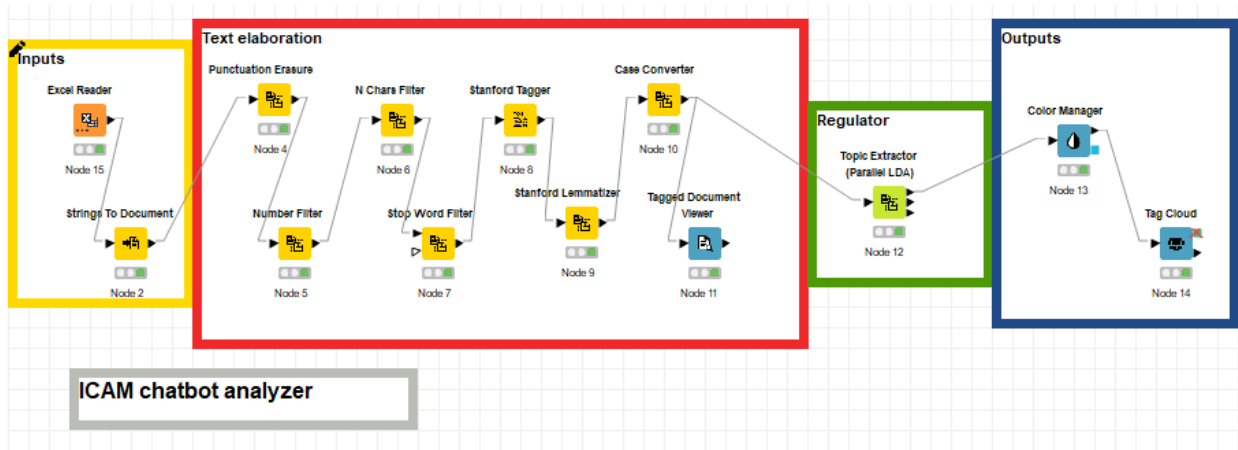


Figure 8. ICAM Chatbot Analyzer.

KNIME Nodes for Tag Cloud	
Macro Process	Nodes
Inputs	Is the initial process of the workflow that is oriented to import data. The macro-process input is based on two different nodes that are “ <i>Excel Reader</i> ” and “ <i>Strings to Document</i> ”.
Text Elaboration	Is the main macro-process of the workflow oriented to analyze the tag cloud. The text elaboration macro-process is based on the following nodes that are “ <i>The Punctuation Erasure</i> ”, “ <i>Number Filter</i> ”, “ <i>N Chars Filter</i> ”, “ <i>Stop Word Filter</i> ”, “ <i>Stanford Tagger</i> ”, “ <i>Stanford Lemmatizer</i> ”, “ <i>Case Converter</i> ”, “ <i>Tagged Document Viewer</i> ”

Regulator	<i>IS "Topic Extractor-Parallel LDA"</i>
Outputs	<i>"Color Manager", "Tag Cloud"</i>

Table 2. KNIME Nodes for Tag Cloud.

The weight effectively can be considered as the frequency of the single word in a distribution of words. In our analysis, we have chosen to put together two or three words in a single expression to obtain a graphical representation in the Tag Cloud that can be useful for the reader. In effect if we would have chosen not to merge two or three words in a single expression, we would have not had the possibility to plot a Tag Cloud effectively readable.

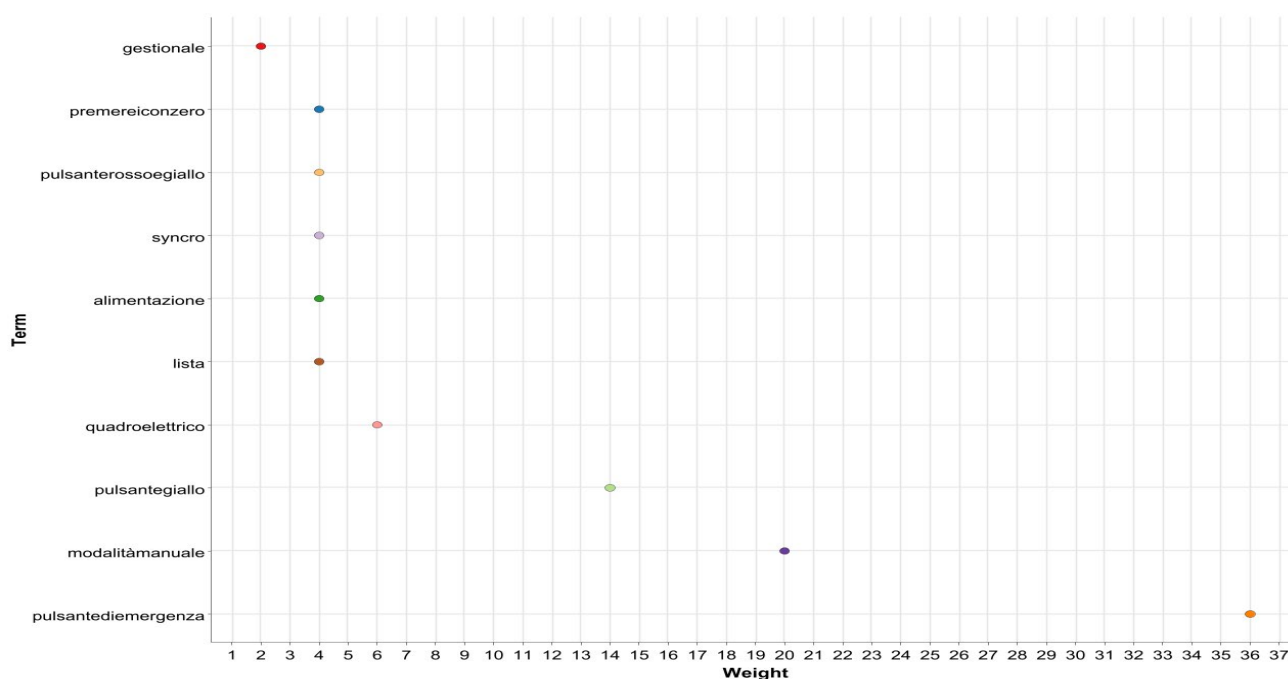


Figure 9. The relationship between Term and Weight.

After the analysis we can effectively plot the tag cloud using four different algorithms. We have created an order of algorithms based on their ability to give the maximum possible representation to the maximum number of words in the tag cloud. Specifically, we have ordered the various functions that can be used in KNIME to plot the Tag Cloud based on their ability to promote readability not only for the words with a greater frequency but also for words with minor frequency. The order we obtained is as follows:

1. *Logarithmic*: by choosing this function in KNIME it is possible to give a representation to a greater number of words even. In effect, as it is shown in Figure 10, the logarithmic representation of the Tag Cloud can plot each word with a good visual result. It is possible to recognize the more frequent word, but also less represented words are visible even if with a smaller dimension.
2. *Square Root*: is at the second rank in the sense of algorithm that give the maximum representation to the maximum number of words. In effect, as it is possible to see in the Figure 10, the more

frequent word is well represented, but also other words with minor frequency are visible even if with a dimension smaller than the Logarithmic function.

3. *Linear*: is at the third rank in terms of visual representation of words in the Tag Cloud as shown in the Figure 10. Specifically, the most frequent word i.e., “*PulsanteDiEmergenza*” is well visible at the center of the Tag Cloud, while the other words that have lower frequency are less visible and someone is not readable at all. The level of readability of the Linear algorithm applied to the Tag Cloud is lower than the Square Root and the Logarithmic option.
4. *Exponential*: is at the fourth rank in terms of readability of the maximum number of words. In effect in this case only the most represented word is readable, while the other words are represented in the background, and are essentially not useful for the reader. For this reason, the use of the option Exponential can be considered at the lowest rank in respect to Logarithmic, Square Root and Linear.

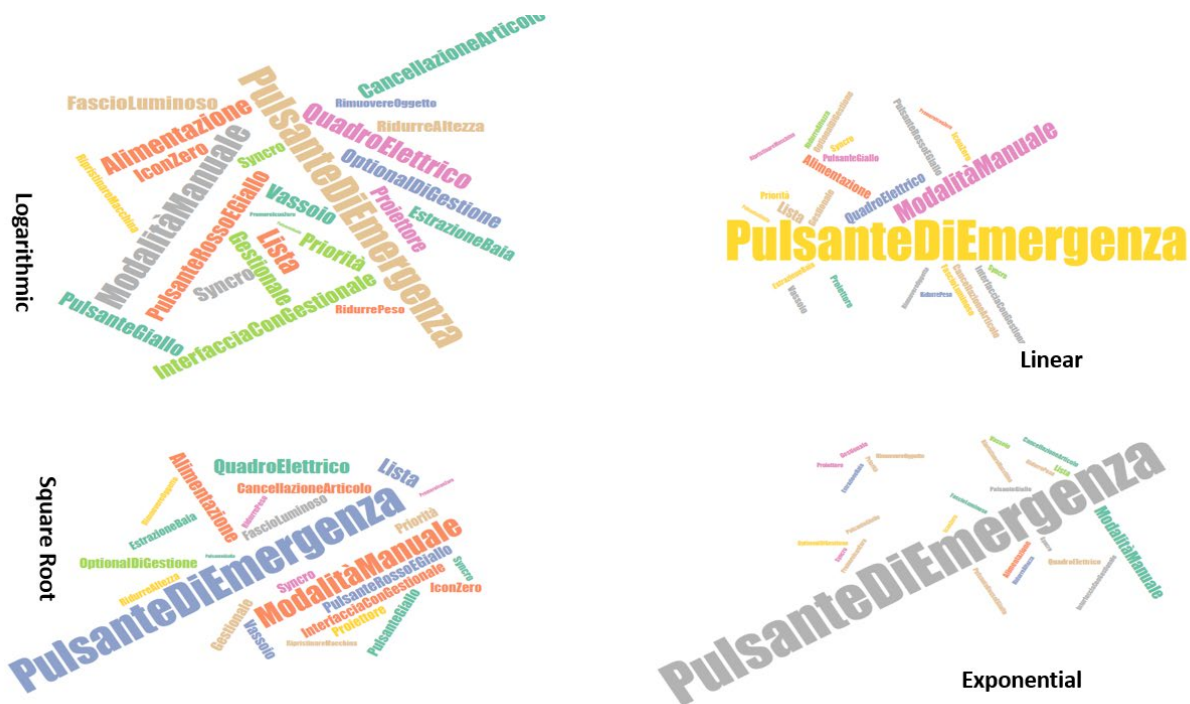


Figure 10. Various typologies of cloud tags.

As we can see from the Figure 10 it is possible to create the following order of algorithm for Tag Cloud i.e.: *Logarithmic* > *Square Root* > *Linear* > *Exponential*. It should be clear that this order is based on the idea of readability i.e., the possibility to offer the maximum representation in the Tag Cloud also for words that have a lower frequency in the distribution. In the third table are summarize the tag cloud experimental results:

Frequencies, Meaning and Customer Care Aspects of the Main Tag Clouds
--

Tag Cloud	Frequency-Occurrences	Meaning	Customer Care Aspect
<i>PulsanteDiEmergenza</i>	36	<i>Emergency Button</i>	The usage of the most frequently used word that is “ <i>PulsanteDiEmergenza</i> ” (Emergency Button), that in the first step of hierarchical clustering is associated with various clusters, in the second step of the process model can suggest the necessity to intervene to improve the level of security of the automated warehouses
<i>ModalitàManuale</i>	20	<i>Manual Mode</i>	The manual functions can be unclear: the manual modality must be improved. This is useful to find new manual functions matching with the automated system.
<i>PulsanteGiallo</i>	14	<i>Yellow Button</i>	The recurrence of Yellow Button shows the necessity of the customer to solve questions that are associated to security. The security aspect is a key element for marketing function-customer sensitivity in security.
<i>QuadroElettrico</i>	6	<i>Electrical Cabinet</i>	The electrical components are relevant for the customer that must related with the automated warehouse. The recurrence of the tag cloud shows the necessity to improve the electrical cabinet to augment customer care.

<i>Lista</i>	4	<i>List</i>	The command list may be unclear, and the improvement of instructions required.
<i>Alimentazione</i>	4	<i>Power Supply</i>	To optimize the procedure to start up the warehouse after the blackout.
<i>Syncro</i>	4	<i>Synchronization</i>	Automated system should be clearer and easier to use.
<i>Pulsanterossoegiallo</i>	4	<i>Button Red and Yellow</i>	Doors open error.
<i>Premereiconzero</i>	4	<i>Press i-Button with zero</i>	Block i with zero.
<i>Gestionale</i>	2	<i>Management</i>	Exchange data by means of an interfacing procedure.

Table 3. Frequencies, Meaning and Customer Care Aspects of the Main Tag Clouds.

5. Conclusions

In this article we have analyzed the role of chatbots in the context of industrial services. Specifically, we propose the case study of ICAM, a firm that operates in the context of automatic warehouses. The usage of AI-Chatbot in the context of industrial services is very popular. In our analysis we use the Business Process Model and Notation-BPMN to investigate the change from the “*AS-IS*” to “*TO-BE*” with a particular application to “*Sales Customer Service*” and “*Technical Support Service*”. The BPMN models put in evidence the various steps and passages of the proposed processes to verify the efficacy of the solution. BPMN models are very useful in the context of process design and process application. We show how the change from the traditional methodology of customer assistance to the AI-Chatbot improves the ability of the firm to generate value for its customers, improving customer care, customer assistance and customer loyalty. Specifically, we have found that the introduction of the AI-Chatbot has positive impacts on Customer Process Management or on Product Quality Management Process with significant impact on customer relationship management. In our analysis the passage from a customer assistance offered by phone and to the AI-Chatbot may generate a relevant impact in terms of relationship marketing with a positive effect on market reputation and brand value.

In the following part, a two-step process model is used to connect technological improvements and relationship marketing in chatbot assistance: the first step provides the hierarchical clustering able to classify questions and answers through Latent Dirichlet Allocation -LDA- algorithm, and the second one executes the Tag Cloud representing the visual representation of more frequent words contained in the experimental dataset. Using a set of questions and answers generated in the relation between the AI-Chatbot and the customer, we determine the structure of the demand based on the hierarchical clustering. Using KNIME, we find that there are 24 hierarchical clusters of which 14 with a singular element, 6 with 2 elements, 1 with 3 elements, 1 with 18 elements, 1 with 10 elements, 1 with 7 elements. Finally, we realize the representation of a Tag Cloud optimized with four different functions i.e., linear, logarithmic, square root, exponential. We order the different typologies of algorithms based on their ability to offer a clear representation not only of the expressions with greater frequency but also of expressions with minor frequency. We prefer readability of the entire set of the expressions present in the dataset. Based on this assumption we find that the logarithmic option offers the best solution for Tag Cloud, and we obtain the following order: *Logarithmic*>*Square Root*> *Linear*> *Exponential*.

Acknowledgement

All the applications have been deployed by a unique IT collaborative framework developed within the Smart District 4.0 Project: the Italian Fondo per la Crescita Sostenibile, Bando “*Agenda Digitale*”, D.M. Oct. 15th, 2014, funded by “*Ministero dello Sviluppo Economico*”. This is an initiative funded with the contribution of the Italian Ministry of Economic Development aiming to sustain the digitization process of the Italian SMEs. Authors thank to the partner Noovle for the collaboration provided during the work development. The proposed results are used to verify the usability of the data on the platform.

Authors

Alessandro Massaro. Professor Alessandro Massaro (ING/INF/01, FIS/01, FIS/03) carried out scientific research at the Polytechnic University of Marche, at CNR, and at Italian Institute of Technology (IIT) as Team Leader by activating laboratories for nanocomposite sensors for industrial robotics. He is in MIUR register as scientific expert in competitive Industrial Research and social development. He was the head of the Research and Development section and scientific director of MIUR Research Institute Dyrecta Lab S.r.l. Member of the International Scientific Committee of Measurers IMEKO and IEEE Senior member, he received an award from the National Council of Engineers as Best Engineer of Italy 2018 (Top Young Engineer 2018). He is currently researcher at



LUM Enterprise S.r.l., and professor at LUM University Libera Università Mediterranea "Giuseppe Degennaro".

Nicola Magaletti, Business development manager graduated in mechanical engineering with over 30 years of work experience in structured companies offering IT solutions and consultancy services, working on the management of innovation processes and in the launch of new business initiatives . Since 2018 he has been part of the Lum Enterprise team as Operational Manager and Technical-Scientific Manager of the “Smart District 4.0” R&D project of which the company is the lead partner.



Gabriele Cosoli, Senior IT Specialist and Solution architect with a degree in Computer Science and over five years of previous experience, specialized in the analysis and design of ICT solutions in various application areas and technological frameworks. Certified on "*Machine Learning by Stanford University on Coursera*" Master on "*Agile and Digital Project Management - Advanced Course*" at 24ORE Business School.

Giardinelli Vito. Project Manager with a degree in Economics and Management (2016, Lum University, Casamassima, Italy) and a Master in Business Strategy and Entrepreneurship (2017, Sda Bocconi, Milan, Italy). Experience: (i) 2014, staff audit, KPMG S.p.A.; (ii) from 2017 to 2020, project manager, Facile.it (Italian price comparison web-site); (iii) from 2021, Lum Enterprise, business developer.



Angelo Leogrande is assistant professor at the LUM University-Giuseppe Degennaro in Casamassima-Bari, where he is also researcher for the LUM Enterprise S.r.l. a spin off oriented to develop digitalization services for SMEs. He worked at Dyrecta Lab Research Institute, a research institute officially recognized by the Italian Minister of University and Research-MIUR, where he acquired professional hard skills on the economic consequences of Industry 4.0, Big Data, Machine Learning and Artificial Intelligence. His research interests include Cooperative Banking, Business Ethics, Innovation Technology, Knowledge, and R&D.

6. Figure Index

Figure 1. Customer process management and product quality management in the case of AI-Chatbot. 5

Figure 2. The organizational change between AS-IS and TO-BE in the case of AI-Chatbot for the case study of ICAM.	6
Figure 3. Technical Support Service.....	Errore. Il segnalibro non è definito.
Figure 4. Sales Customer Service.	7
Figure 5. Workflow of hierarchical clustering.	11
Figure 6. The composition of clusters. On the left side with the indication of Row there are the elements of the clusters. For each row there is a correspondent set of questions-answers.	12
Figure 7. The composition of clusters. On the left side with the indication of Row there are the elements of the clusters. For each row there is a correspondent set of questions-answers.	12
Figure 8. ICAM Chatbot Analyzer.	13
Figure 9. The relationship between Term and Weight.....	14
Figure 10. Various typologies of cloud tags.	15
Figure 11. BPMN model to describe the chatbot functionality of ICAM.	26
Figure 12. BPMN of the Sales Customer Service.	28

7. Bibliography

- Al Islami, M. T. F., Barakbah, A. R. & Harsono, T., 2020. Interactive Applied Graph Chatbot with Semantic Recognition. *2020 International Electronics Symposium (IES)*, pp. 557-564.
- Ashok, M., Ramasamy, K., Snehitha, G. & Keerthi, S. R., 2021 . A Systematic Survey of Cognitive Chatbots in Personalized Learning Framework. *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 241-245.
- Behera, R. K., Bala, P. K. & Ray, A., 2021. Cognitive Chatbot for personalised contextual customer service: Behind the scene and beyond the hype. *Information Systems Frontiers*, pp. 1-21.
- Chao, M. H., Trappey, A. J. & Wu, C. T., 2021. Emerging Technologies of Natural Language-Enabled Chatbots: A Review and Trend Forecast Using Intelligent Ontology Extraction and Patent Analytics. *Complexity*.
- Chou, Y. C., Chao, C. Y. & Yu, H. Y., 2019. A Résumé Evaluation System Based on Text Mining. *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pp. 052-057.
- Chung, M., Ko, E., Joung, H. & Kim, S. J., 2020. Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, Volume 117, pp. 587-595.
- Dharani, M. et al., 2020. Interactive Transport Enquiry with AI Chatbot. *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1(IEEE), pp. 1271-1276.
- Dharaniya, R., Vijayalakshmi, K., Tejasree, R. & Naveena, P., 2020. Survey on interactive chatbot. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* , 8(IV).
- Dodero, J. R., Flaviani, F. & Alarcón, S., 2020. Analysis on Formality Classification in Conversational Bots. *Octava Conferencia Nacional de Computación, Informática y Sistemas / CoNCISa*, pp. 67-71.

- Følstad, A. & Taylor, C., 2021. Investigating the user experience of customer service chatbot interaction: a framework for qualitative analysis of chatbot dialogues. *Quality and User Experience*, 6(1), pp. 1-17.
- Hien, H. T. et al., 2018. Intelligent assistants in higher-education environments: the FIT-EBot, a chatbot for administrative and learning support. *Proceedings of the ninth international symposium on information and communication technology*, pp. 69-76.
- Ikemoto, Y., Asawavetvutt, V., Kuwabara, K. & Huang, H. H., 2018. Conversation strategy of a chatbot for interactive recommendations. *Asian Conference on Intelligent Information and Database Systems*, Springer(Cham), pp. 117-126.
- Kowatsch, T. et al., 2017. Text-based healthcare chatbots supporting patient and health professional teams: preliminary results of a randomized controlled trial on child.
- Kushwaha, A. K., Kumar, P. & Kar, A. K., 2021. What impacts customer experience for B2B enterprises on using AI-enabled chatbots? Insights from Big data analytics. *Industrial Marketing Management*, Volume 98, pp. 207-221.
- Landim, A. R. D. B. et al., 2021. Chatbot design approaches for fashion E-commerce: an interdisciplinary review. *International Journal of Fashion Design, Technology and Education* .
- Lommatzsch, A., 2018. A next generation chatbot-framework for the public administration. *International Conference on Innovations for Community Services*, pp. 127-141.
- Luo, B., Lau, R. Y., Li, C. & Si, Y. W., 2021. A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p. 1434.
- Madhu, D. et al., 2017. A novel approach for medical assistance using trained chatbot. *2017 international conference on inventive communication and computational technologies (ICICCT)*, pp. 243-246.
- Maeda, H., Saiki, S., Nakamura, M. & Yasuda, K., 2019. Recording Daily Health Status with Chatbot on Mobile Phone-A Preliminary Study. *2019 Twelfth International Conference on Mobile Computing and Ubiquitous Network (ICMU)*, pp. 1-6.
- Mantravadi, S., Jansson, A. D. & Møller, C., 2020. User-Friendly MES Interfaces: Recommendations for an AI-Based Chatbot Assistance in Industry 4.0 Shop Floors. *12th Asian Conference on Intelligent Information and Database Systems*.
- Massaro, A., 2021. Electronic in Advanced Research Industry: From Industry 4.0 to Industry 5.0 Advances. *Wiley/IEEE*, p. ISBN.
- Massaro, A. & Galiano, A., 2020. Re-engineering process in a food factory: an overview of technologies and approaches for the design of pasta production processes. *Production & Manufacturing Research*, 8(1), pp. 80-100.
- Massaro, A., Gargaro, M., Birardi, G. & Galiano, A., 2021. Combined CNN and LSTM Models for Robotic Financial Guidance. *Informatica*, 4(32), pp. 2-20, .

- Massaro, A., Maritati, V. & Galiano, A., 2018. Automated Self-learning Chatbot Initially Build as a FAQs Database Information Retrieval System: Multi-level and Intelligent Universal Virtual Front-office Implementing Neural Network. *Informatica*, 42(4).
- Massaro, A., Meuli, G., Savino, N. & Galiano, A., 2020. Voice Analysis Rehabilitation Platform based on LSTM Algorithm. *International Journal of Telemedicine and Clinical Practices*, Volume 1.
- Massaro, A., Vitti, V., Galiano, A. & Morelli, A., 2019. Business intelligence improved by data mining algorithms and big data systems: an overview of different tools applied in industrial research. *Computer Science and Information Technology*, 7(1), pp. 1-21.
- Mathew, R. B., Varghese, S., Joy, S. E. & Alex, S. S., 2019. Chatbot for disease prediction and treatment recommendation using machine learning. *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 851-856.
- Murad, D. F. et al., 2019. Towards smart LMS to improve learning outcomes students using LenoBot with natural language processing. *2019 6th International Conference on Information Technology Computer and Electrical Engineering (ICITACEE)*, pp. 1-6.
- Nagarhalli, T. P., Vaze, V. & Rana, N. K., 2020. A Review of Current Trends in the Development of Chatbot Systems. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Volume IEEE, pp. 706-710.
- Omogbe, N. A. et al., 2020. Text Messaging-Based Medical Diagnosis Using Natural Language Processing and Fuzzy Logic. *Journal of Healthcare Engineering*.
- Pantano, E. & Pizzi, G., 2020. Forecasting artificial intelligence on online customer assistance: Evidence from chatbot patents analysis. *Journal of Retailing and Consumer Services*, Issue 102096, p. 55.
- Paul, A., Haque Latif, A., Amin Adnan, F. & Rahman, R. M., 2019. Focused domain contextual AI chatbot framework for resource poor languages. *Journal of Information and Telecommunication*, 3(2), pp. 248-269.
- Pizzi, G., Scarpi, D. & Pantano, E., 2021. Artificial intelligence and the new forms of interaction: Who has the control when interacting with a chatbot?. *Journal of Business Research*, Volume 129, pp. 878-890.
- Prasad, V. A. & Ranjith, R., 2020. Intelligent Chatbot for Lab Security and Automation. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Volume IEEE, pp. 1-4.
- Przegalinska, A. et al., 2019. In bot we trust: A new methodology of chatbot performance measures. *Business Horizons*, 62(6), pp. 785-797.
- Qaffas, A. A., 2019. Improvement of Chatbots semantics using wit. ai and word sequence kernel: Education Chatbot as a case study. *International Journal of Modern Education and Computer Science*, 11(16).
- Sari, A. C. et al., 2020. Chatbot Developments in The Business World. *Advances in Science, Technology and Engineering Systems Journal*, 5(6), pp. 627-635.

- Setiawan, B. M., Zulhawati, Z., Maitri, A. L. & Sutopo, J., 2019. Chatbot Services at Educational Institutions with Customer Relationship Management. *Journal of International Conference Proceedings (JICP)*, 2(1), p. 17.
- Sheth, A., Yip, H. Y. & Shekarpour, S., 2019. Extending patient-chatbot experience with internet-of-things and background knowledge: case studies with healthcare applications. *IEEE Intelligent Systems*, 4(34), pp. 24-30.
- Siangchin, N. & Samanchuen, T., 2019. Chatbot implementation for ICD-10 recommendation system. *2019 International Conference on Engineering, Science, and Industrial Applications (ICESI)*, Volume IEEE, pp. 1-6.
- Sidaoui, K., Jaakkola, M. & Burton, J., 2020. AI feel you: customer experience assessment via chatbot interviews. *Journal of Service Management*.
- Tjiptomongsoguno, A. R. W. et al., 2020. Medical Chatbot Techniques: A Review. *Proceedings of the Computational Methods in Systems and Software*, Springer(Cham), pp. 346-356.
- Yorita, A., Egerton, S., Chan, C. & Kubota, N., 2020. Chatbot for Peer Support Realization based on Mutual Care. *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Volume IEEE, pp. 1601-1606.

8. Appendix

The BPMN of Technical Support Service. The Business Process Management and Notation-BPMN of the ICAM AI-Chatbot is shown in the figure 11. The BPMN is composed of four different pools that are interrelated.

The entire process starts with the platform smart district 4.0 that collects information about customers. Subsequently the platform acquires details of the products to assist the client. The platform sends a message to the client with the indication of the problem encountered. The client exposes the problem to the platform SD40 with a message. The platform SD4.0 analyzed the question and verified the necessity to open a ticket. At this point the platform evaluate if the ticket is already opened and two options are feasible:

- *Yes, the ticket was already opened:* in this case the platform re-opens the existing ticket;
- *No, the ticket was not already opened:* in this case the platform acquire data for the opening of the ticket and effectively open a new ticket;

At this point the platform Smart District 4.0 sends a ticket opening notification to the operator and two news possibilities are feasible since the proposed request was already known:

- *Yes, the problem was already known:* in this case the platform acquires information, sends the solution, closes the ticket, stores the conversation, sends the conversation at the list of

distribution, and then closes the sub-process. In sending the solution to the customer two alternatives are feasible based on whether the assistance was provided or not:

- *Yes, the assistance was provided:* in this case the customer analyzes the answer. And two new states are possible since the question was solved:
 - *Yes, the problem was solved:* the customer closes the procedure of assistance and later the process ends;
 - *No, the problem was not solved:* the customer improves details at the problem encountered and later closes the ticket.
- *No, the assistance was not provided:* the customer stores are taking charge and later the process ends.
- *No, the problem wasn't already known:* the SD40 platform verifies the presence of economic coverage. At this point two different stages are feasible base on the presence of economic coverage:
 - *Yes, there is economic coverage:* the platform checks the availability of the operator. At this point two different stages are feasible base on the availability of the operator:
 - *No, the operator is not available:* the platform acquires contacts with the customer, communicates taking charge, stores the conversation, sends the conversion to the distribution list, and the process ends.
 - *Yes, the operator is available:* the platform sends the problem to the Icam operator. The Icam operator analyzes the problem. At this point two different stages are feasible since it is necessary to carry out a maintenance intervention.
 - *No, the maintenance intervention is not necessary:* in this case the Icam operator sends the solution to the smart district 4.0 platform. Later the platform submits a solution, stores information, updates knowledge base, stores conversation, sends conversation to distribution list. But during the phase of the submission of the solution to the customer two new states become feasible:
 - *Yes, the assistance was provided:* in this case the customer analyzes the answer. And two new states are possible since the question was solved:
 - *Yes, the problem was solved:* the customer closes the procedure of assistance and later the process ends;
 - *No, the problem was not solved:* the customer improves details at the problem encountered and later closes the ticket.
 - *No, the assistance was not provided:* the customer stores are taking charge and later the process ends.

- *Yes, the maintenance intervention is necessary:* at this point two different stages are feasible based on the following question “*Is it advisable to outsource the maintenance intervention?*”:
- *No, the outsource intervention is not advisable:* Notify date of assistance, store appointment, close the ticket on SageX3 and the process ends.
- *Yes, the outsource intervention is advisable:* Notify date of assistance, store appointment, close ticket on SAGE X3. The external operator analyzes the question, stores an appointment, and later the process ends.
- *No, there isn't economic coverage:* the platform updates the ticket on SAGEX3, activates the sales customer service and finally the process ends.

- *No, there is not an activation of the macro process:* in this case the platform asks for product and service information. In this case there are two options: the platform can either send a message to the customer that later enters information about the service / product and sends a message to the platform to acquire information on the product / service, or alternatively can acquire information relative to the product/service.
- In the following steps the platform acquires information relative to the product/service and later realizes the following stages: identifies product / service, opens business opportunity, and notifies the business opportunity to the commercial operator. At this point two different possibilities are feasible based on “*is the product / service known?*”:
- *Yes, the product/service is known:* in this case the platform retrieves information, sends draft commercial offers, updates business opportunity, notifies the assistance taken in charge, stores conversation, sends conversation to distribution list. And later the process ends. The platform during the phase of sending commercial offer drafts also sends a commercial offer to the ICAM commercial operator that operates the following steps: analyzes request, identifies product / service, sends commercial offer, updates business opportunity. During the sending of the commercial offer the ICAM operator also can send a message to the platform SD4.0 that realizes the following steps: sends the commercial offer, stores the information, updates the knowledge base, stores the conversation, sends the conversation to the distribution list, and ends the process. But during the phase of sending the commercial offer it is possible that the platform sends a message to the customers. At this point two different options are feasible based on the following question: “*Has the commercial offer been received?*”:
- *Yes, the commercial offer has been received:* in this case the customer analyzes the answer and later two new stages become feasible based on the following question “*Has the offer been accepted?*”:
- *Yes, the offer has been accepted:* in this case the customer sends an order and later the process ends.
- *No, the offer has not been accepted:* in this case two new steps are feasible based on the following question “*Was the service slot communication received?*”:
- *Yes, the service slot communication was received:* in this case the customer stores the assistance slot and later the process ends.
- *No, the service slot communication was not received:* in this case the process ends.
- *No, the commercial offer has not been received:* in this case the process ends.

- *No, the product/service is not known:* in this case the platform realizes the following steps: agrees assistance slot, updates business opportunity, reports assistance slot, stores assistance slot, stores conversation, sends conversation to distribution list and ends the process.

The BPMN of the Sales Customer Service is represented in Figure 12.

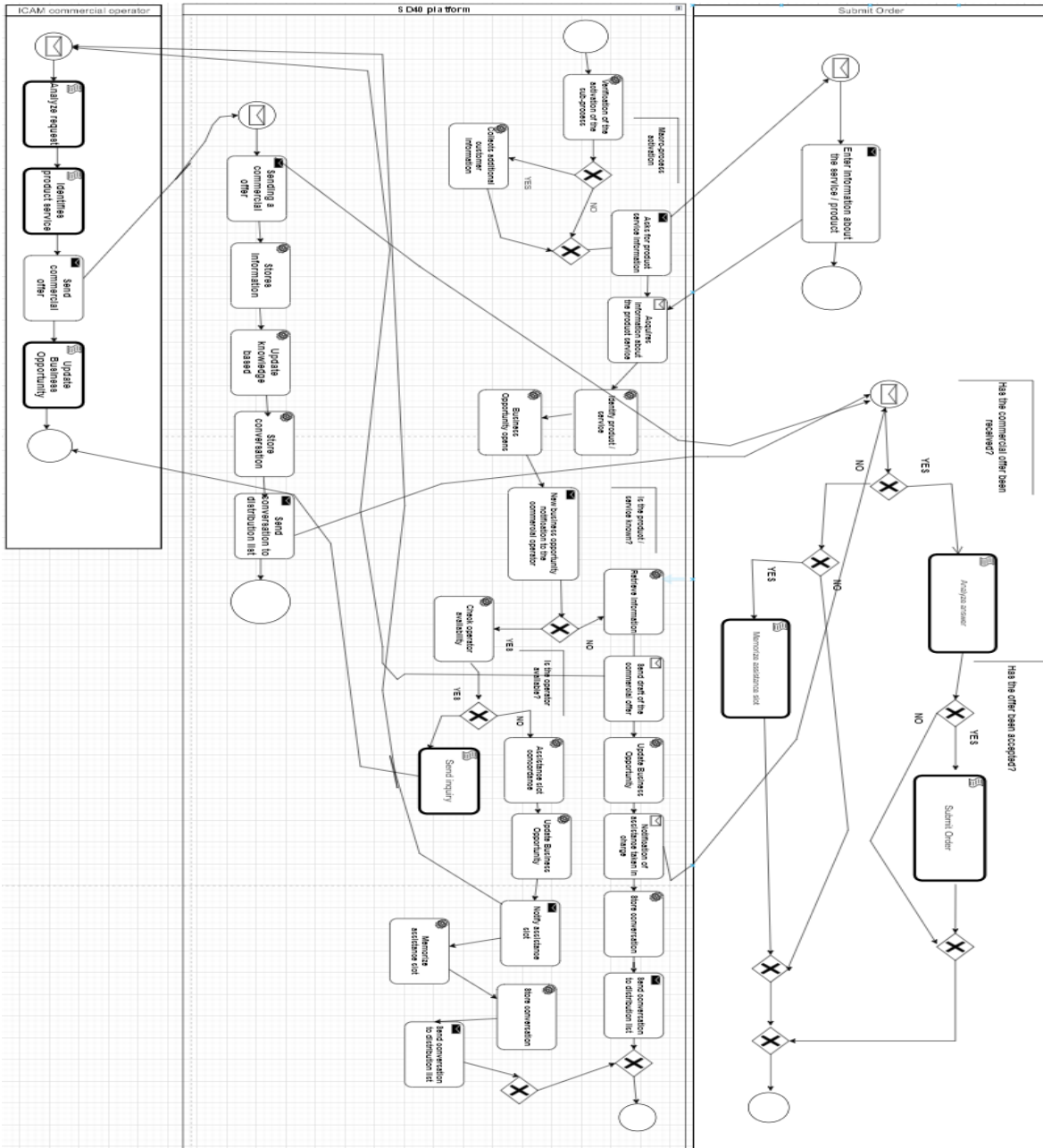


Figure 12. BPMN of the Sales Customer Service.