CMS Optimization with Deep Learning Techniques

Alberto Schiaffino[2], Matteo Reina[2], Ricardo A.Matamoros A.[1][0000−0002−1957−2530], Francesco Epifania[1], Francesco Ruggeri[1][0000−0002−5827−2565], Ignazio Maria Castrignano[1][0000−0003−1123−2520], and Luca Marconi[1][0000−0002−0236−6159]

[1] Social Things SRL, Italy
{ricardo.matamoros, francesco.epifania, francesco.ruggeri, ignazio.castrignano, luca.marconi}@socialthingum.com,
[2] Engitel S.p.a., Italy
{alberto.schiaffino, matteo.reina}@engitel.com

Abstract. This study stems from the observation of the growing demand, interest and incredible potential offered by Artificial Intelligence and Machine Learning techniques in the fields of content website or e-commerce.
In particular, we will study the Content Management System (CMS), a system capable of creating and maintaining websites, forums and applications for customers. These systems have an intrinsic predisposition to provide services based on Machine Learning (ML), thus being able to add to the solutions offered by the most innovative ML systems.
These include Sentiment Analysis, the Recommender System and the more innovative Chatbot .
Integrating them into a website or web platform very often leads to significantly improved results for the user-experience offered to the user.
This study aims to analyse the new artificial intelligence techniques and how to implement them in a company. Reporting also a case of Study based on a real agency.

Keywords: Machine Learning · Content Management Systems · Artificial Intelligence · Sentiment Analysis · Natural Language Processing

1 Context description

The objective of the study is to create a usable Platform-as-a-Service (PaaS) extension that makes available and integrates different Machine Learning services so that the company can offer these solutions to its customers.

To this end, an infrastructure will be created to show an interface through which CMS can request different ML services as needed. In particular, CMS will only have to indicate which specific ML module it wants to use within a site (whether a Sentiment Analysis module [2], items recommendations [1], ...) and some parameters; the infrastructure will interpret the request and select the correct module to be queried.
For some of these, there are already various solutions offered by providers such as Microsoft (Azure), Google (Google Cloud Platform) and Amazon (AWS); in these cases, the services will be studied and tested to see which ones meet the requirements.
Where the offered solutions are not adequate, the necessary ML models will be built from scratch, starting from the state of the art and improving it with more recent techniques. In particular, given the central importance in websites and the absence of an adequate service, the item recommendation system will be studied.

The realisation of the project involves the design and development of two components: the architectural infrastructure and the ML-based services.

2 Technology implementation

By analysing the time required for the various phases of the implementation, we have arrived at significant estimates. The first macro-phase is aimed at the realisation of the infrastructure that integrates the services and makes them easily usable.
The specific sub-phases, with their corresponding time estimates, are:
Numbered (ordered) lists are easy to create:

1. Infrastructure design and development (1 month)
2. Design and development of procedures for data life cycle management (0.5 month)
3. Design and development of the communication interfaces with the Content Management System (0.5 months)

The second macro-phase is aimed at the implementation of basic services with ML. The specific sub-phases are:

1. Study of the state of the art of existing ML services (0.5 months)
2. Integration of ML models already offered by external services (1.5 months)
3. State of the art of Recommender Systems (RS) (0.5 months)
4. Model evaluation and testing (1.0 months)

The aim is to create a working prototype with at least TRL 5, to verify and validate the effectiveness of the technologies applied to the reference context before moving on to the next phase.

3 State-of-the-art innovation

The implementation of these technologies within a company produces first and foremost Product Innovation, as it proposes a radical enhancement of the proposing company's current solution. This will allow the company to open up to new markets, to meet demands that it could not previously satisfy.
But the main value of these technologies is that they will enable important simplifications and reduction of implementation times in the business processes of customers using the Content Management System.
The CMS enables:

- more timely, effective and efficient data collection;
- a complete processing of data and information through a wide variety of tools;
- overall integration between them
- the provision of new services that did not previously exist

Therefore, the implementation will affect: the enhancement of management activities, optimisation and monitoring of business processes through artificial intelligence and Machine Learning techniques. In fact, the multiple synergies that it stimulates within the processes, allow a restructuring of the interaction between technology and organization, thus digital transformation in a broad sense.

3.1 Unique and complete system

The artificial intelligence-based services offered by some providers, such as Microsoft, Google and Amazon, are 'black box', they are neither unified nor integrated, nor are the solutions complete with all services.
The proposed infrastructure, on the other hand, is aimed at a comprehensive and dynamic management of the whole range of tools and services and the necessary data.
In particular, the infrastructure will deal with:

- management of the data life cycle
- creation and updating of models
- communication between the infrastructure and the system

The architecture of the infrastructure will make it possible to implement the full spectrum of services required, either by developing them from scratch or by exploiting those already offered by the providers, thus enabling interchangeability and the choice of the most performing ones.

3.2 Study and implementation of new ML models

Particular attention was paid to different Recommender System models. It starts with a state-of-the-art study of the most popular algorithms currently implemented to try out more recent techniques based on neural networks. In this way it is possible to test and evaluate different solutions, describing which ones and in which contexts they provide better results.
Some services are related to all or some of the following fields:

- Recommender systems: item recommendations customised for the user
- Sentiment Analysis: extraction of opinion from a text [9][20]
- Content Moderator: collection of user contributions (comments, generic text and image content) and their automatic moderation to filter out inappropriate or out of scope content [3][4]
- Text analysis for SEO [12][19]
- Chatbot or advanced search to refine the content searched for the user [7] [14] [16]

In this way, a significant improvement in the state of the art will be possible, which will have a significant impact on the competitive advantage of the company and of the customers who adopt its system.

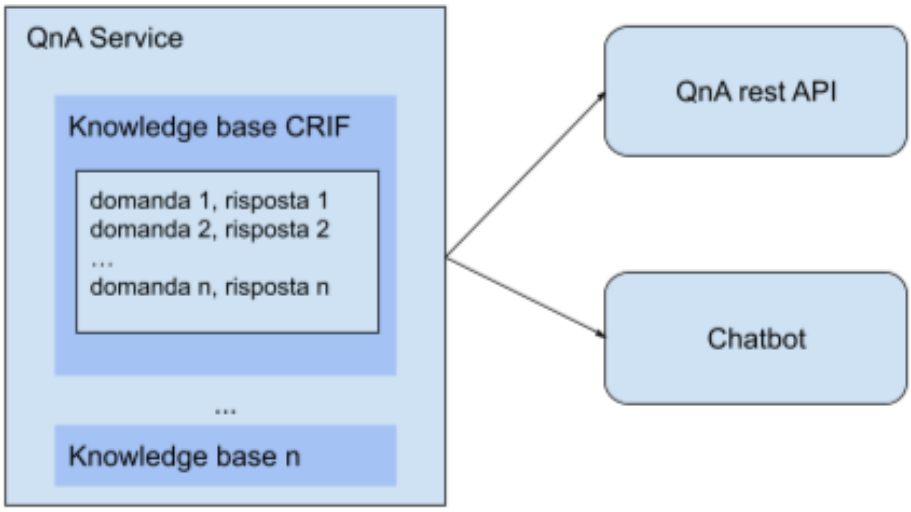

Fig. 1. In the figure we see at a very high level the link between the various services used in this context.

4 Economic impact

4.1 How to finance it

The breakdown of costs according to the actions envisaged in the project is indicatively as follows:

1. Activities to identify the Artificial Intelligence services useful to complete the CMS platform. (6% of costs)
2. Design and development of the infrastructure with CMS interface. (30% of costs)
3. Integration and implementation of basic services with ML.(52% of costs)
4. User Interface validation activities.(12% of costs)

4.2 Impact on Enterprise

A second macro-result that can be highlighted, from the point of view of the effects on productivity and on the increase in revenues, is that relating to the management of data, including unstructured data: the possibility for the infrastructure to jointly manage heterogeneous and also unstructured data and obtain significant insights, thanks also to economies of scale and scope and knowledge, allows in the medium term a reduction in organisational costs for data processing and knowledge management in the broadest sense [13] [15]. These potential productivity gains from optimised knowledge management [17] and the parallel decreases in organisational costs are then also transferred to the organisational structure of customers, who see significantly improved performance from a data analytics perspective.
And that's not all: the numerous possibilities for creating and selling customised services, as well as customised data and information based on the needs of customers, means that differentiated pricing logics can be configured, making information available on request according to the logic of the "Data as a Service" paradigm.

4.3 Impact on the Environment

As far as product innovation is concerned, the proposed system is characterised first and foremost by the potential environmental benefits that can be obtained by integrating services into a single infrastructure. This entails numerous energy-saving factors deriving from the univocal and omni-comprehensive management of the entire data life cycle.
As can be seen from the study carried out applying this technology to e-learning courses. The resources recommended by the system will be improved in quality and quantity through external sources as they have done in their work by Garg et al. [8].
In particular, the user will analyse the data by means of a single architectural structure integrating the different services, which are variously complementary

from the point of view of the obtainable insides: consequently, the use of the energy resources of the Content Management System will be structured in a progressively better way according to the reorganisation of the production processes stimulated adaptively by the use of the platform itself. In this way, the content provided to the user will dynamically adapt to his needs in a more specific and precise manner and navigation within the system will be facilitated, with significant savings in time and energy.
Following this, the improved performance of the system will progressively stimulate further savings, thanks to Process Intelligence, through the highlighting of organisational bottlenecks at each stage of the process and data life cycle. Ultimately, the resulting environmental, social and economic benefits will be perceptible from the point of view of both the company and the customers themselves, and therefore of the entire ecosystem that uses the services generated by the CMS.

In addition, since the chosen architecture is that of a Platform-as-a-Service, all processing linked to the Machine Learning algorithms will be carried out on Cloud machines optimised for this type of processing, reducing consumption in terms of energy used and heat produced, as we seen in the work of Rajabion, Lila, et al.. [18]
Finally, during the entire life cycle of the project, the choice of the Cloud implies further important benefits for the entire system from the point of view of environmental impact and in terms of saving on the costs of energy for local data and materials for server hardware.

5 Case of Study

Over the last few years Engitel has developed its own Content Management System (CMS) called Spin&GO (SG) with which it creates and maintains websites, forums and applications for its customers. Spin&GO is increasingly a Made in Italy excellence. Engitel has already participated in the project called SPAC3, as part of the projects funded by the Lombardy region as Bando Smart Cities 2014 and the Easy 4.0 project.
The system currently used by Engitel, Spin&Go, does not offer any solution based on Artificial Intelligence; the advantages that can be achieved through these techniques are by now well known to all and in this case they mainly concern the quality of the experience provided to the user while navigating the site or using the application: from the suggestion of personalised content based on the user, to the automation of content insertion procedures, to the automatic recognition of multimedia content.

The SPIN&GO BRAINY project therefore opens the way to the use of AI applications and environments within the CMS and the Engitel processes.
The use of such services therefore leads to a higher quality of the products provided by Engitel, consequently with an increase in turnover for the company.

Engitel turned to Social Things which has all the complementary competences in the Artificial Intelligence field necessary to realise SPIN&GO BRAINY.

Therefore, the SPIN&GO BRAINY project allows the opening of the proposing company to new markets and new clients (who could currently use other providers), through the provision of Machine Learning services: this significantly expands the offer and enriches the value proposition of Engitel itself.

On the one hand, it is a unique system especially at the Italian level and unified of different synergic or complementary ML services, which allows for an easy integration within websites and applications.
Therefore, the SPIN&GO BRAINY project fully falls within the macro-area of Intelligent Business Process Management, as defined by Gartner [22]: that is, the enhancement of management activities, optimisation and monitoring of business processes by means of Artificial Intelligence and Machine Learning techniques.

Studying the economic impact on the company we note that:
Engitel's revenue over the last 4 years has been roughly as follows:

- Year 2017: 3,795,448€.
- Year 2018: 4.207.526€.
- Year 2019: 3.716.874€.
- Year 2020: 3.743.917€.

The Net Profit after tax (in thousands of Euros) amounted to:

- Year 2017: 685,508€.
- Year 2018: 656,868€.
- Year 2019: 430,321€.
- Year 2020: 451,196€.

From the above figures, Engitel is able to tackle this project with its own means. The financial and budget parameters recorded in the last fours years and the forecasts make this possible, as the investment of approximately €40,000 represents approximately 7% of the average net profit of the last three years and the company is not in debt.
The price differentiation resulting from customisation and the user-centred offer will enable Engitel and its customers to progressively increase revenues.
In the case of Engitel, it will be necessary to recruit highly specialised technical staff to manage the enhanced CMS: therefore, figures will be needed to deal with system integration and system administration, for the overall management of the platform, but also data scientists and full-stack developers to design specific ML modules. It is estimated that there will be an increase of about 10 people in three years.

In the case of Social Things, an additional two engineers specialised in data science and AI will be needed to design the algorithms and modules, while an

additional full-stack developer will take care of the infrastructure design.

In fact, Engitel currently manages each process and the resulting services for specific customers in an individual and totally differentiated way. For each web service offered, there are specific resources allocated, resulting in energy redundancies due to waste and lack of synergies.
By means of the SPIN&GO BRAINY platform, from the point of view of production processes, the optimised management of knowledge deriving from the integration of synergic services will instead allow an overall saving in terms of energy and production efficiency.

6 Implementation details

In this work we plan to use Microsoft cloud tools for the delivery of the ML services provided. As we can observe below, in the figure 2, the architecture and data flow of this proposed work had a simple and intuitive representation, specifically we will use:

- Azure Data Factory and Azure Synapse Analytics: automating the extraction, loading and transformation (ELT) workflow
- Azure Databricks and Azure SQL Server Analysis Services: for analytics and batch data management.
- Azure API management: for publishing APIs for internal and external consumers

The importance of cloud services lies in the improvement of software scalability as it allows the optimisation of continuous integration and release processes of functionalities, both in the creation and update phases. In particular, the services for the hierarchical orchestration of functionalities introduce greater flexibility in terms of structural independence by providing the possibility to efficiently implement the configuration of microservices for the CMS.
The 4 proposed services will be used through cloud APIs according to a set of characteristics:

- on-demand and self-service
- high accessibility and flexibility
- dynamic grouping of resources by adapting the allocation to physical storage and computing devices according to requests
- service measurement and monitoring and control to ensure transparency of service use.

The choice of Microsoft Azure cloud infrastructure relates to a number of technological and operational aspects, but also to recognitions received in the API sphere. In fact, Microsoft was listed as a leader in Gartner's 2021 Magic Quadrant report for Comprehensive API Lifecycle Management.

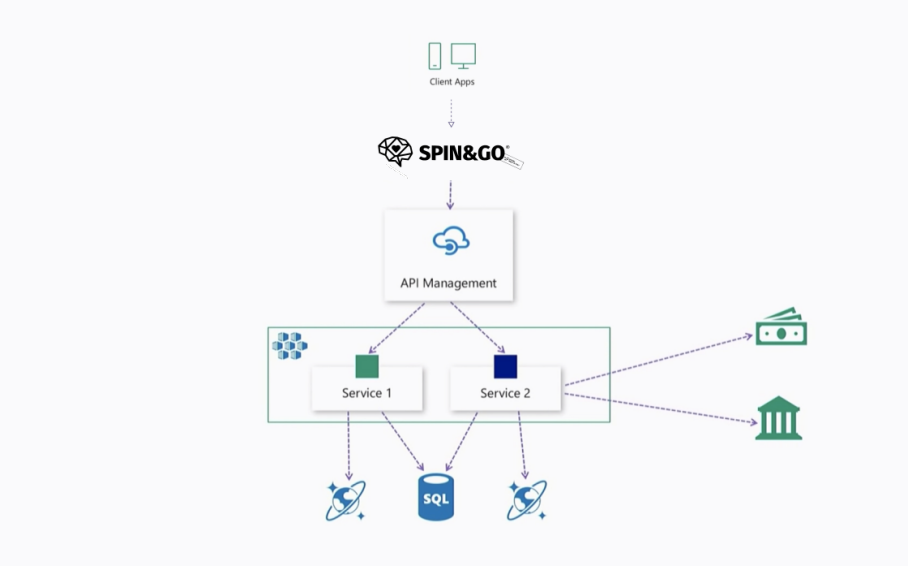

Fig. 2. Azure API architecture

7 Conclusions and Future Developments

In this study we have analysed the implementation of new technologies based on Machine Learning to Content Management Systems.
The problem addressed in this work, besides being complex because it concerns a vast domain of active search, presents other problems related to the need to apply it on a real case, i.e. on a company that offers such services.
Has we can see in the work of Bonhard et al. [6] it is possible to significantly improve the quality of recommendations by exploiting the knowledge derived from users profile and their social network pages.
The case study analysed in this paper shows how Engitel through its CMS called SPIN&GO offers the indicated web services and by creating instead a PaaS that integrates new Machine Learning technologies it is able to guarantee the enhancement of management activities, optimisation and monitoring of business processes through Artificial Intelligence and Machine Learning techniques.
The actual state-of-the-art envisages the use of deep learning technique in the area of Natural Language Processing, in particular the best performing model now are based on Recurrent Neural Networks and/or reinforcement learning as in the work of Subramanyam et al. [21] in which are described the the model based on the self-supervised learning and transfer learning [10].
Another research line propose as approach to NLP models based on attentional mechanism as introduced in the paper of Hu and Dichao [5].

References

1. Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).
2. Hussein, Doaa Mohey El-Din Mohamed. "A survey on sentiment analysis challenges." Journal of King Saud University-Engineering Sciences 30.4 (2018): 330-338.
3. Gorwa, Robert, Reuben Binns, and Christian Katzenbach. "Algorithmic content moderation: Technical and political challenges in the automation of platform governance." Big Data & Society 7.1 (2020): 2053951719897945.
4. Gillespie, Tarleton. "Content moderation, AI, and the question of scale." Big Data & Society 7.2 (2020): 2053951720943234.
5. Bonhard, Philip, and Martina Angela Sasse. "'Knowing me, knowing you'—Using profiles and social networking to improve recommender systems." BT Technology Journal 24.3 (2006): 84-98.
6. Subramanyam Kalyan, Katikapalli, Ajit Rajasekharan, and Sivanesan Sangeetha. "AMMUS: A Survey of Transformer-based Pretrained Models in Natural Language Processing." arXiv e-prints (2021): arXiv-2108.
7. Nuruzzaman, Mohammad, and Omar Khadeer Hussain. "A survey on chatbot implementation in customer service industry through deep neural networks." 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE). IEEE, 2018.
8. Garg, Vishal, and Ritu Tiwari. "Hybrid massive open online course (MOOC) recommendation system using machine learning." International Conference on Recent Trends in Engineering, Science & Technology-(ICRTEST 2016). IET, 2016.
9. Young, Tom, et al.. "Recent trends in deep learning based natural language processing." ieee Computational intelligenCe magazine 13.3 (2018): 55-75.
10. Hu, Dichao. "An introductory survey on attention mechanisms in NLP problems." Proceedings of SAI Intelligent Systems Conference. Springer, Cham, 2019.
11. Alyafeai, Zaid, Maged Saeed AlShaibani, and Irfan Ahmad. "A survey on transfer learning in natural language processing." arXiv preprint arXiv:2007.04239 (2020).
12. Vedantam, Vamsi Krishna. "The Survey: Advances in Natural Language Processing using Deep Learning." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.4 (2021): 1035-1040.
13. Zaslavsky, Arkady, Charith Perera, and Dimitrios Georgakopoulos. "Sensing as a service and big data." arXiv preprint arXiv:1301.0159 (2013).
14. Bala, K., et al.. "Chat-bot for college management system using AI." International Research Journal of Engineering and Technology 4.11 (2017): 2030-2033.
15. O'Leary, Daniel E. "Using AI in knowledge management: Knowledge bases and ontologies." IEEE Intelligent Systems and Their Applications 13.3 (1998): 34-39.
16. Lalwani, Tarun, et al.. "Implementation of a Chatbot System using AI and NLP." International Journal of Innovative Research in Computer Science & Technology (IJIRCST) Volume-6, Issue-3 (2018).
17. Mendes Rodrigues, Eduarda, and Natasa Milic-Frayling. "Socializing or knowledge sharing? Characterizing social intent in community question answering." Proceedings of the 18th ACM conference on Information and knowledge management. 2009.
18. Rajabion, Lila, et al.. "A new model for evaluating the effect of cloud computing on the e-learning development." Journal of Workplace Learning (2019).
19. Gjorgjevska, Emilija, and Georgina Mirceva. "Content Engineering for State-of-the-art SEO Digital Strategies by Using NLP and ML." 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). IEEE, 2021.

20. Lee, Gichang, et al.. "Sentiment classification with word localization based on weakly supervised learning with a convolutional neural network." Knowledge-Based Systems 152 (2018): 70-82.
21. Lee, Joohong, Sangwoo Seo, and Yong Suk Choi. "Semantic relation classification via bidirectional lstm networks with entity-aware attention using latent entity typing." Symmetry 11.6 (2019): 785.
22. Wang, John, et al.. "Cloud Computing, Intelligent Business Process Management and Artificial Intelligence." international journal of data analysis and information systems 9 (2017): 1-12.