COMBINING DATA ENRICHMENT AND PREDICTIVE ANALYTICS FOR DIGITAL MARKETING CAMPAIGN IMPACT OPTIMIZATION

Marta López[1], Fernando Perales[1][0000−0003−3304−0925], Cynthia Parrondo[1], Flavio De Paoli[2][0000−0002−5047−7371], Matteo Palmonari[2][0000−0002−1801−5118], and Nikola Tulechki[3][0000−0002−7318−1637]

[1] JOT INTERNET Media, Madrid 28020, Spain
marta.lopez@jot-im.com, fernando.perales@jot-im.com, cynthia.parrondo@jot-im.com
http://www.jot-im.com
[2] DISCo, University of Milano-Bicocca, Milan, Italy
flavio.depaoli@unimib.it, matteo.palmonari@unimib.it
[3] Ontotext, Sofia, Bulgaria
nikola.tulechki@ontotext.com
http://www.ontotext.com

Abstract. In the dynamic realm of digital marketing, fine-tuning campaign strategies through analytical data introspection becomes a central focus for competitive advantage. JOT is at the forefront of this innovation with two pioneering initiatives: Intelligent Keyword Enrichment Analytics Service for launching highly efficient digital marketing campaigns) and marketing data enrichment with weather data to identify correlations. The first one leverages advanced semantic analysis and machine learning algorithms to optimize keyword research and monetization, dynamically adapting to market changes and consumer behaviour. The second approach complements this by enriching campaign data with external variables such as weather conditions and significant events, with the aim of uncovering hidden patterns and correlations that inform campaign optimization. This work presents a dual approach to revolutionizing digital marketing effectiveness: the semantic power enhances keyword potential, while the contextual enrichment provides a deeper understanding of the environmental factors that impact consumer behaviour. Together, they signify a paradigm shift in targeted advertising and strategic planning, enabling JOT to refine its campaigns for maximum engagement and ROI.

Keywords: Semantic Web · Data Integration · Web Technologies · Digital Marketing Optimization · Semantic Data Enrichment · Event-Influenced Marketing · Climate Data Analysis · Predictive Analytics · Semantic Pattern Recognition · Web Traffic Analysis.

1 Introduction

In recent years, digital marketing has undergone unprecedented changes in data value chain due to the need to connect efficiently and cost-effectively with specific audiences. The increased importance of context has led to keyword research and impact optimization becoming essential tools in determining the success of advertising campaigns and their associated profitability. Against this backdrop, investment in digital marketing is expected to continue to increase, reflecting the growing importance of targeted and personalized advertising strategies [1]. However, the sheer amount of information and the complexity of digital consumer behaviour pose specific challenges, especially in terms of identifying legitimate traffic and optimizing campaigns for real audiences.

In this environment, this work has emerged as a pioneer with the development of two innovative solutions for data analytics and keyword classification, InMa, and for data preparation and linking, enRichMyData (EMD) toolkit.

InMa stands out for its use of advanced semantic analysis and machine learning algorithms, which enable a deeper and more nuanced understanding of the intent behind user searches. Not only does this approach improve the relevance of keywords selected for campaigns, but it also facilitates agile adaptation to changing search trends and consumer preferences. InMa's ability to predict the language and category of keywords becomes a strategic advantage, allowing for more effective targeting and personalization of campaigns.

EMD, on the other hand, brings a new innovative dimension to campaign analysis by integrating contextual data such as major events and weather variables. Enriching the data helps companies understand more clearly how these external factors can affect user behavior and, therefore, the effectiveness of advertising campaigns. By linking this data to campaign performance metrics, EMD can detect patterns and correlations that were previously hidden, providing valuable information to improve advertising strategies.

The synergy between InMa and EMD represents a significant advance in the development of more advanced and effective solutions to digital marketing challenges. By marrying the power of semantic analysis [2] with the richness of contextual data, these tools provide a complete view of the relationship between keywords, content, and the consumer. In addition to improving the accuracy and relevance of advertising campaigns, this integrated approach also creates new opportunities for strategic innovation in digital marketing.

This study explores how the two approaches combined, offer a comprehensive methodology for addressing and overcoming today's digital marketing challenges. This work seeks to illuminate the potential of these tools to transform digital marketing practices, ensuring that campaigns not only reach their target audience but also resonate with them through a detailed analysis of their capabilities.

2 Motivation and Background

Today's digital marketing faces a key challenge due to the increasing sophistication of automation algorithms used by leading digital platforms such as Google. In recent years, moreover, advertising campaign management has been revolutionised by these 'smart bidding' systems, which offer automated methods to optimise advertising performance. However, the automatic and opaque nature of these algorithms can restrict the ability of companies to directly influence campaign optimisation, reducing the extent to which they can intervene to improve results.

In this context, accurate keyword selection and correct contextualisation of advertising campaigns becomes crucial. JOT, with its massive operation involving the management of 5 billion active keywords and the generation of more than 1,000,000 clicks per day in 234 countries and 15 languages, is faced with the challenge of preserving the effectiveness and efficiency of its campaigns in this automation environment. The need for advanced analytics and optimisation tools is underlined by adding another layer of complexity - the ability to adapt to diverse cultural and linguistic contexts.

It is a fact that the scale of data that JOT works with has surpassed the capacity of any manual approach, making it imperative to adopt automated solutions. To address this need, JOT has created several algorithms that are based on optimisation rules to effectively manage campaigns at large scale. Maintaining competitiveness and efficiency in an increasingly complex digital marketing environment has essentially depended on this automation. However, JOT's primary focus on extensive keyword deployment has raised concerns about the digital footprint and effectiveness of campaigns. Too many keywords can decrease the effectiveness of campaigns and unnecessarily increase the digital footprint, which is increasingly important in a digital world that values sustainability and resource efficiency.

JOT is responding to these challenges through the development of the InMa and EMD approaches. EMD recognises the importance of context in optimising campaigns and enriches advertising data with relevant information about events and weather conditions. Based on this and the need for smarter and more adaptive keyword selection to better understand the intentions behind user searches, InMa focuses on predicting changes in market trends and making our campaigns more efficient by choosing how best to distribute our keywords.

The origin of the EMD model comes from the urgent need to identify whether there are seasonal keywords [3] in the broad range of our advertising campaigns categories. Although it is commonly assumed that certain keywords are seasonal [4], empirical evidence of this phenomenon has been difficult to demonstrate due to the large amount of data required and the lack of a detailed tracking of campaigns performance that would allow for more specific analyses, such as at city level, rather than national or international observations. This lack of concrete data and the unknown about the potential impact of weather conditions in the user behaviour as another seasonal factor led us to pose two fundamental questions for a deeper understanding of our advertising strategies. The affirmation of these hypotheses implies a radical revision of our current strategies for launching advertising campaigns, focusing on a more dynamic and contextualised approach that takes into account seasonality and environmental variables as critical factors in their planning and execution.

InMa, meanwhile, stems from the realisation that manual keyword optimisation is not feasible given the scale of JOT's operations. With hundreds of millions of active keywords and the need to adapt to market trends in real time, it is imperative to have a system that not only handles keyword selection efficiently but also predicts their potential for success. InMa responds to this challenge with an approach based on algorithms [5] and advanced data analysis.

In sum, these innovative tools combined promise a significant advantage in a market dominated by automated bidding algorithms, where traditional optimisation is constrained. By understanding and using context effectively, and selecting keywords with greater precision [6] and relevance, JOT seeks to improve not only the quality of traffic directed to clients' websites, but also to maximise profit in an increasingly automated and competitive advertising landscape [7].

3 Technical Framework: The Data Flow

In this section, we will address in detail how data flows, showing how action helps uniquely in the optimization and efficiency of digital marketing campaigns. It is structured around four strategic questions designed to drive campaign efficiency and a deep understanding of campaigns:

1. Does the specific time at which a keyword is triggered affect its performance?
2. Do user behavior patterns vary according to prevailing weather conditions?
3. How can the accuracy of keyword selection within a large dataset be increased?
4. What methods can effectively predict the future relevance of a keyword?

It is crucial to answer these questions to move beyond traditional marketing strategies and move toward a system that incorporates machine learning and predictive analytics. The goal is to find the perfect keywords aligned to market trends and weather forecast to adjust campaigns accordingly. We have also seen that using advanced techniques such as approximate keyword research in semantic external databases [11] can achieve more accurate and relevant targeting in our campaigns, further optimizing the effectiveness of our digital marketing strategy.

The diagram in Fig. 1 illustrates the complete integration of the workflow, which combines automated marketing data collection with the advanced capabilities of the EMD and InMa platforms. The strategy starts with the collection of marketing performance data from advertising platforms such as Google Ads, highlighting the importance of capturing information efficiently and accurately, thus generating our preliminary dataset. In our case and taking into account the large volume of data, the use of cloud tools such as Google Cloud and BigQuery becomes very relevant.

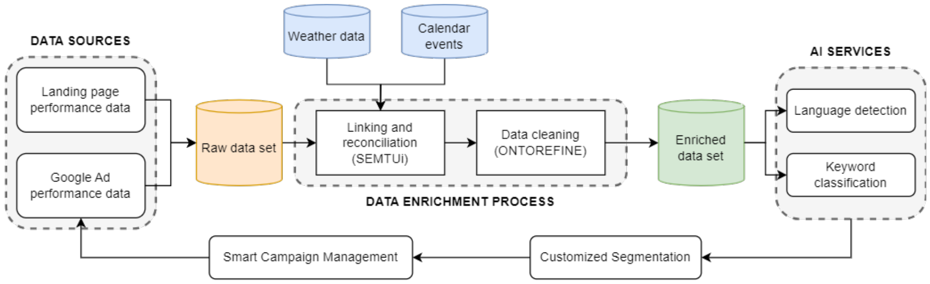


Fig. 1. Implemented workflow enabling data enrichment and processing with AI services for customized segmentation.

After the collection of the economic and performance data generated by the different campaigns, the flow is directed to EMD pipeline, where the data is initially enriched with external information such as weather data with the help of the SemTUI tool (see section 4.1). This is followed by the corresponding cleaning of the data, which is essential for the next steps, and in this case we do this using OntoRefine (see section 4.2). The reliability and accuracy of both processes critical, as the information generated results in an enriched database, whose information will later be translated into optimisation actions and campaign adjustments to achieve better performance and efficiency. Importantly, marketing and weather data are carefully linked by using the date, latitude and longitude to provide a detailed and complete context on the conditions in which advertising campaigns operate. The addition of this dimension allows for analysis at different levels, taking into account geographic seasonality and preparedness for adverse weather conditions such as snowfall or storms, which can vary considerably from region to region.

As the next workflow cycle we find artificial intelligence tools focused on analysing enriched marketing data. These include language detection using FastText embedding model and keyword classification. But there are more tools in development that will act in a complementary way, introducing seasonality analysis and regression-based predictive models, all with the aim of performing deep analysis of our data. The introduction of underlying artificial intelligence will personalise and increase the effectiveness of the targeted market segment. Our accuracy and relevance in advertising depends on this discerning capability, which tailors campaigns to the cultural and linguistic subtleties of each specific audience. In addition, through the use of deep neural networks, it simultaneously performs keyword ranking. The ranking is not fixed, but changes and develops through seasonality analysis and the use of regression-based predictive models, providing more detailed knowledge about how certain keywords and categories vary in their effectiveness over time and in different contexts. Therefore, intelligent processing is not only limited to retrospective analysis, but is proactive in enabling JOT to anticipate the future relevance of keywords with predictive models. InMa offers personalised and dynamic targeting, enabling JOT marketers to adjust and fine-tune their campaigns with unprecedented precision by marrying these facets of artificial intelligence with the user interface. In this way, it establishes itself as the foundation of a digital marketing strategy that recognises and adjusts to continuously evolving consumer behaviour and market trends, ensuring that each campaign not only achieves its objective, but also maximises ROI with qualified and engaged traffic.

This is where JOT's internal processes come into play. Through in-depth examination, relationships are identified and predictive models are developed to support optimisation decisions. Campaigns are adapted and refined based on this analysis, allowing for agile response to contextual dynamics and strategic adaptation. For example, by including variables such as daylight hours or categorisation and the impact of local events on search trends, campaigns are more accurately aligned with current conditions. In this way, a keyword that would normally be unprofitable can become a valuable opportunity when contextualised within an appropriate timeframe, such as school holidays or festive periods. Therefore, we could summarise that the first core of the JOT workflow is made up of three phases: collection, enrichment and application. This allows JOT to not only analyse performance from a cost-effectiveness perspective, but also to take a holistic approach that considers the context to make more informed and strategic decisions.

In summary, the integrated workflow highlights how JOT uses advanced artificial intelligence to not only collect and enrich data, but also to interpret and apply it in effective and efficient advertising campaigns. JOT emphasises its commitment to the forefront of digital advertising through a coordinated and synergistic approach, with each component and development contributing to the achievement of strategic objectives in a market dominated by automation and personalisation. The EMD model is established as an analytical tool created to integrate detailed records and InMa improves efficiency and effectiveness through intelligent processing. Our main focus when developing campaigns is Search Engine Marketing (SEM), with the goal of increasing visibility on search engine results pages (SERPs), especially on Google. Through this approach, we improve campaigns to drive qualified traffic.

4 Components Development

This section presents the technical advances and key developments that form the basis of our integrated analysis and optimisation system. We will look in detail at how the structure of our system facilitates an efficient and consistent workflow, focusing on the progress made in data processing and analysis. Thanks to this holistic approach, we are able to overcome the challenges of collecting and contextualising data, constantly optimising campaigns to achieve and exceed our performance and efficiency targets.

In the project context, we have the advantage of using advanced tools that improve and streamline our work processes.

4.1 SemTUI

SemTUI [8][9] is a tool that focuses on supporting users in data enrichment through reconciliation and linking of mentions in tabular data, as well as extending such data with information taken from external sources. It provides a simple user interface that enables even non-expert users to exploit semantics to discover useful services that can be combined in a pipeline process to achieve a given goal. In our case, the objective of augmenting the source marketing data with meteorological data needs identifying weather services capable of providing the required data, along with the information required to retrieve it. In our case, this involves reconciling locations and associating latitude and longitude coordinates that, along with time information, enable for querying weather services.

SemTUI's reconciliation and linking capability relies on an extensible set of services, either based on Knowledge Graphs (KGs) such as Wikidata, or datasets such as HERE[1] and Open Street Map[2] (OSM). A key advantage of using SemTUI [10] is the capability to visually review the outcomes, enabling access to and potential correction of the results (see Fig. 2). The extension capability allow users to integrate the table content with meteorological information from services such as OpenMeteo[3] and the European Centre for Medium-Range Weather Forecasts (ECMWF)[4] APIs (see Fig. 3). The manual identification of steps in the enrichment process (e.g., reconciliation with a geocoding service that returns coordinates, followed by extension services to retrieve meteorological data) on a sample of the full dataset can lead to the definition of a pipeline that can be automatically executed on the entire dataset, and replicated weekly, to prepare the data for marketing analysis. The process not only establishes a direct

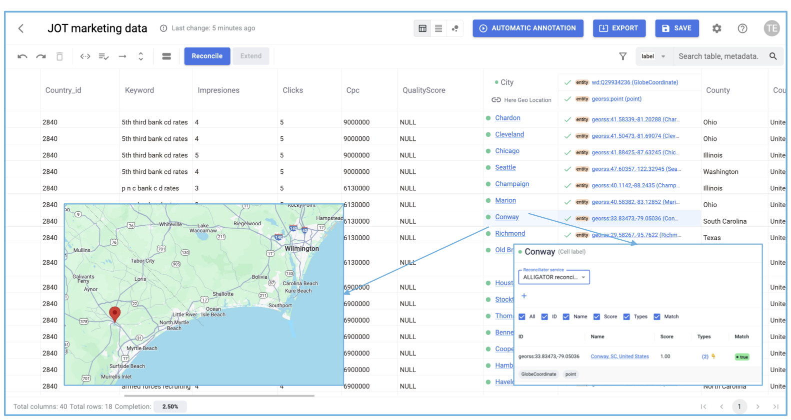
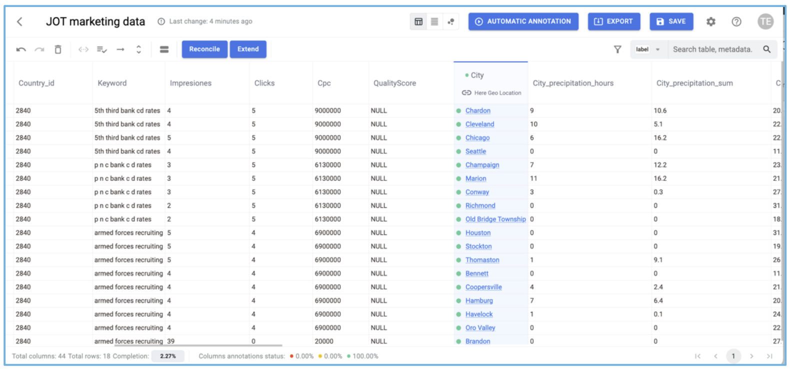


Fig. 2. Visual reconciliation and review with SemTUI.

[1] https://developer.here.com/develop/rest-apis
[2] https://publicapis.io/open-street-map-api
[3] https://open-meteo.com/en/docs
[4] https://www.ecmwf.int/en/computing/software/ecmwf-web-api

connection between weather information and campaign performance in different geographic locations, but also automates the collection of essential data to improve campaigns. The outcome of the enrichment process is expected to lead to deeper insights that impact marketing strategy, improve decision-making and personalise campaigns to maximise performance.



Country_id	Keyword	Impresiones	Clicks	Cpc	QualityScore	City	City_precipitation_hours	City_precipitation_sum	C
2840	5th third bank cd rates	4	5	9000000	NULL	Chardon	9	10.6	20
2840	5th third bank cd rates	4	5	9000000	NULL	Cleveland	10	5.1	22
2840	5th third bank cd rates	5	5	9000000	NULL	Chicago	6	16.2	22
2840	5th third bank cd rates	4	5	9000000	NULL	Seattle	0	0	11
2840	p n c bank c d rates	3	5	6130000	NULL	Champaign	7	12.2	23
2840	p n c bank c d rates	3	5	6130000	NULL	Marion	11	16.2	21
2840	p n c bank c d rates	3	5	6130000	NULL	Conway	3	0.3	27
2840	p n c bank c d rates	2	5	6130000	NULL	Richmond	0	0	31
2840	p n c bank c d rates	2	5	6130000	NULL	Old Bridge Township	0	0	18
2840	armed forces recruiting	5	4	6900000	NULL	Houston	0	0	31
2840	armed forces recruiting	5	4	6900000	NULL	Stockton	0	0	19
2840	armed forces recruiting	5	4	6900000	NULL	Thomaston	1	9.1	26
2840	armed forces recruiting	4	4	6900000	NULL	Bennett	0	0	11
2840	armed forces recruiting	4	4	6900000	NULL	Coopersville	4	2.4	21
2840	armed forces recruiting	4	4	6900000	NULL	Hamburg	7	6.4	20
2840	armed forces recruiting	4	4	6900000	NULL	Havelock	1	0.1	24
2840	armed forces recruiting	4	4	6900000	NULL	Oro Valley	0	0	22
2840	armed forces recruiting	39	0	20000	NULL	Brandon	0	0	27

Fig. 3. An example of enriched table.

4.2 Ontotext Refine

Ontotext Refine[5] is an indispensable tool in the data standardisation and cleaning process for JOT, as it executes the necessary transformations in an efficient and automated manner, resulting in the generation of consistent and analytically valuable data. The tool is built on top of the Open Refine[6], to which it adds tabular to RDF mapping and transformation capabilities as well as batch processing. It gives data engineers the ability to precisely define the type of each column, discard redundant information or calculate complex statistical parameters such as quartiles and interquartile range directly during data preparation, thus transcending its function beyond just removing anomalies. These valuable features optimise the team's time and focus on handling missing values and limit settings to detect outliers and transform them. With Ontotext Refine, JOT ensures that the data set is purified not only of erroneous data, but also enriched with calculations that support the detection of meaningful trends and patterns to improve advertising campaigns. And this in a quick and easy way by entering a csv file, after which it will dispaly to the data engineer with a suite of tools which to quickly analyse and transform the data to the required shape (see Fig. 4). In other words, thanks to this tool we can dedicate our efforts directly to the analysis of information, saving time and effort for the analysts. The end

[5] https://www.ontotext.com/products/ontotext-refine/
[6] https://openrefine.org/

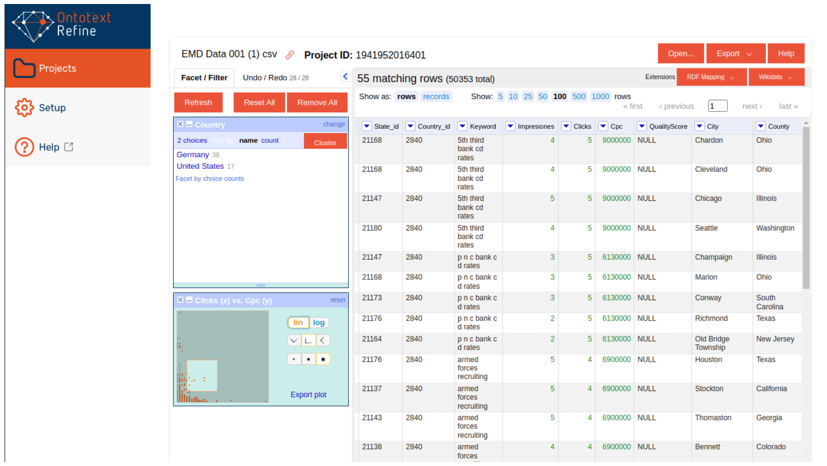


Fig. 4. Ontotext Refine user interface with text and numeric facets

result of this detailed process is a comprehensive database that includes all the elements necessary for a thorough evaluation of our advertising campaigns. A thorough analysis of this rich data set is carried out to uncover possible correlations or patterns between user behaviour and external variables such as events and weather conditions. The integration of this information not only helps us to understand the immediate impact on our campaigns, but also to improve strategic decision making for future advertising initiatives. In short, these tools are part of a methodical, data-driven approach to improving decision making in digital marketing. The meticulous data analysis and subsequent preparation provides unique and valuable insight into how advertising campaigns interact with external factors. As the study moves towards evaluating longer time periods and including a wider range of variables, our expectation is to further strengthen our understanding and strategies for connecting with consumers effectively and efficiently.

4.3 Keywords categorization

In addition, the next fundamental piece is the introduction of artificial intelligence, which, given our daily data volume, is crucial for our business model, as it has become unmanageable manually. The first steps that have been taken to implement this type of methodology in our day-to-day optimisation and uploading of new campaigns are these two main tools:

Language Detection component: The language detection module is a key part of InMa's infrastructure, playing a vital role in adapting campaigns to the linguistic particularities of diverse global markets. This module harnesses the power of the linguistic embeddings generated by fastText (see Fig. 5), an advanced solution for semantic text representation that excels in its ability to work with a large number of languages and its efficiency in handling out-of-vocabulary words and misspellings. fastText's accuracy in language identification [12] becomes a strength, allowing for more refined audience targeting and effective targeting of advertising campaigns. By analysing these vector representations, the system reliably classifies the language of keywords, a step that could ensure cultural resonance and relevance of the advertising message. Since language determines the relevance of the target audience and the marketing strategy applied, this identification is critical. Consistent and efficient execution across different development and production environments is ensured by the tool's integration into a Docker environment. **Category classification**: This module

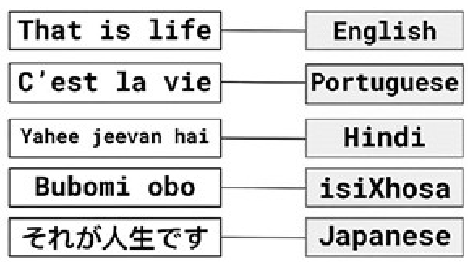


Fig. 5. Examples of Language Detection with fastText

employs a more complicated approach by using a customised deep neural network10 to traverse Google's multi-level hierarchical category structure. We have named this network JOTNet, which is built to handle multi-level classification and thus accurately assign each keyword to its corresponding category within the broad category tree. (see Fig. 6). JOTNet uses keyword embeddings as input and processes them through multiple layers to understand the relationships and characteristics that define each category and subcategory. Due to its deep, hierarchical architecture, JOTNet can take into account the complexity and detail of categories, which facilitates detailed and consistent classification. In fact, it is often able to improve on the original categorisation. The system organises keywords into a complex hierarchical tree consisting of 3182 categories up to seven levels deep. JOTNet is particularly effective at learning optimal paths in the classification tree, allowing for more refined decisions that reflect different levels of abstraction. Nowadays, this type of tool is of great importance, as most of our keywords from keyword research lack this type of information. The absence of them makes it very difficult to select a segmentation for the implementation of different campaigns as we manage daily uploads of thousands of keywords. This has been a step forward in the automation and improvement of the way we work and undoubtedly marks a before and after in the way our company approaches the rest of the available possibilities. Together with these two tools, other tools

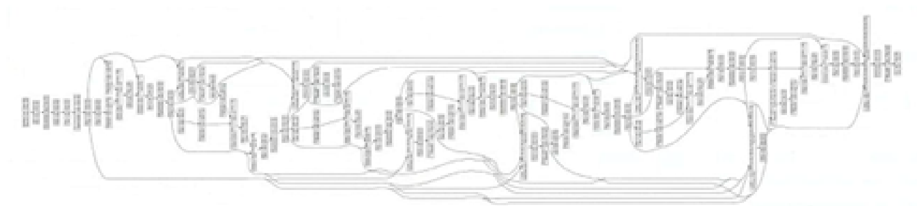


Fig. 6. Implemented workflow enabling data enrichment and processing with AI services for customized segmentation.

are being developed in parallel, aimed at studying seasonality, another great unknown in our field of study. The aim is to implement different predictive models that will further improve the efficiency of our advertising campaigns. The objective is a much more effective segmentation.

5 Evaluation and Results

These initiatives are currently in an early evaluation phase, during which pilot tests give offered first evaluation of how the systems perform and how effective the developed systems are. This section documents the preliminary results obtained and the knowledge acquired during the initial tests, as well as to establish the metrics that will guide future evaluations.

The evaluation of data enrichment has revealed important information during the initial testing phase. Ontotext Refine has been used in the core component to process marketing data, clean, and structure datasets previously enriched with detailed information on events and weather conditions. It was identified that 31% of the values in 'quality_score' were missing during the tests, so they were imputed using the median in order to keep the data distribution intact and preserve the general characteristics of the original set.

Feature engineering was carried out to encode categorical variables and create new metrics like 'daylight_hours', based on the difference between sunrise and sunset. EMD has been able to identify patterns that were previously not considered (see Fig. 7)), such as the influence of daylight hours on consumer behavior. In this case, the tests have been with only two isolated months of 2023, June and December, so we have mostly found intrinsic relationships like the relationship between temperatures and the month.

Responding to the first two main questions posed, we can say:

1. **Does the specific moment when a keyword is triggered affect its performance?** The development of EMD incorporates a temporal analysis that allows us to explore seasonal patterns, to answer this. For instance, revealing significant trends for winter campaign planning can be done through the correlation between click-through rates on cold-weather clothing ads and

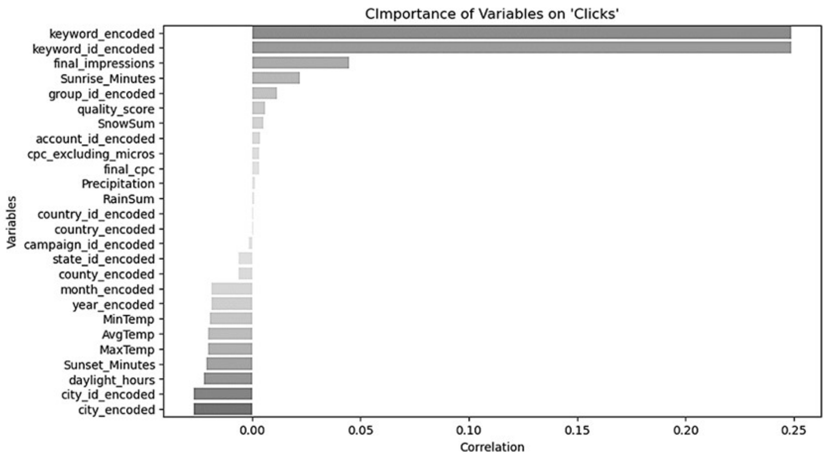


Fig. 7. Implemented workflow enabling data enrichment and processing with AI services for customized segmentation.

low temperatures. Therefore, thanks to the project, we will be able to answer this question.

2. **Do user behavior patterns vary according to the prevailing weather conditions?** Within EMD, it is examined how environmental variables such as precipitation or extreme temperatures influence the performance of keywords by analyzing correlations. This is achieved by evaluating the impact of these variables on key metrics such as clicks and final impressions, which allows campaigns to be adjusted to the actual conditions experienced by users. In other words, thanks to EMD, we will be able to see interesting patterns in this area or discard previously posed hypotheses.

In summary, all of this has managed to create a richer foundation for strategic decision-making in advertising campaigns by analyzing the importance of variables in generating clicks and final impressions, along with these correlations.

On the other hand, predictive models have shown excellent performance, with an average accuracy of 85% in predicting categories thanks to the implementation of the JOTNet neural network. Although managing 3,182 categories distributed across 7 hierarchical levels is complex, the model has managed to maintain a balance in the F1, precision, and recall metrics, which are around 0.65 approximately. We have noticed that the model sometimes chooses categories different from the original labels, which has led in several cases to a more accurate categorization that is more in line with the current market reality (see Table 1) , which explains its values in terms of error. Moreover, this error increases from category level 1, meaning it categorizes differently but not less efficiently. This underscores the effectiveness of the new categorization structure. Regarding language detection, the system has shown outstanding efficiency in accurately identifying the language across a wide variety of keywords (see Ta-

Table 1. . Category prediction for 'how apple care works'

Level	Initial Categorization	Predicted Category
1	Internet&Telecom	Internet&Telecom
2	Unknown	Telephony
3	Unknown	MobilePhones&Accessories
4	Unknown	MobilePhones
5	Unknown	SmartPhones

ble 2) . However, it has been observed that accuracy decreases in the presence of spelling errors and in cases where a keyword spans multiple languages. Although limited, these exceptions provide clear opportunities for continuous improvement of the system.

Table 2. Prediction Results: Keyword by Language and Category

Keyword	Predicted Language	Predicted Category
Horarios autobus 671	es	BusServices
Cuisines en marbre ikea	fr	Kitchen&DiningRoomFurniture
Cheap electronic drumset	en	Drums&Percussion
Kleiner Schreibtisch	de	Office&ComputerDesks

To summarize and address the last two questions posed:

1. How can the accuracy in selecting keywords within an extensive dataset be increased? We can respond to this challenge by analyzing performance history and market trends to select the most promising keywords. This will increase relevance and conversion in campaigns.
2. What are the methods that can effectively predict the future relevance of a keyword? The use of predictive models based on past and present data to anticipate upcoming trends. This enables proactive forecasting of keyword selection to optimize campaigns and capitalize on future market movements and changes in consumer behavior.

This entire system is designed to ensure continuous improvements in algorithms through feedback and results, which, along with the handling of new enriched data, will enrich JOT's marketing campaigns and strategies. On the other hand, identifying spelling errors and keywords in multiple languages are specific areas for future improvements, which will be addressed by implementing more complex contexts and advanced techniques for more accurate disambiguation.

6 Conclusions and Next steps

The preliminary evaluation conducted for the months of July and December 2023 has revealed signs pointing to the influence of climatic factors and events on user responses to our campaigns. Although the correlations found are modest, they represent significant evidence that deserves further investigation. Despite their subtlety, these connections highlight the complex relationship between the environment and consumer behavior, thus emphasizing the importance of a broader database to strengthen our predictive models. The initial findings pave the way for more detailed investigations, with the aim of fully capturing seasonality and other contextual factors affecting the optimization of our advertising campaigns. Therefore, our ability to detect and capitalize on patterns in consumer behavior needs to be refined, as the current methodology and sophistication of our analysis tools have demonstrated. This will allow us to adjust our strategies to an ever-evolving market.

On the other hand, our algorithms have faced similar challenges in the first phase of testing. The language detection and keyword categorization modules have performed as expected, despite the limitation in available data. The results suggest a better understanding of the linguistic context and semantic classification, which could directly influence the keywords segmentation and campaigns update.

According to the tests carried out, the insights gained are valuable and will be used to adapt and improve the system. In the future, tests are planned to collect broader data and achieve results with greater statistical significance. During this process, we have faced multiple challenges, but two main ones can be distinguished: The first problem is caused by typographical errors, which cause the algorithm to misunderstand the word. The second concerns ambiguous words, as we have found examples where it is common to mix words from other languages, and in these cases, the algorithm chooses a word without considering the context. We conclude that it is important to consider not only the keyword itself but also surrounding words or similar words in our predictions to draw more realistic and accurate conclusions. Therefore, the next steps will focus on conducting tests in real environments for each of the tools, with the goal of expanding our database with precise and diverse information that allows for deeper analysis.

Undoubtedly, the advances presented here represent a significant achievement in the field of digital marketing, highlighting not only their promise for JOT but also their transformative power for the industry as a whole. Although the path to total optimization still requires meticulous scrutiny and adjustment, the progress made so far indicates an upward trend toward an unprecedented understanding and interaction with the digital market and its consumer dynamics. Success in digital marketing is defined by the continuous evolution of these components, highlighting a future where analytical precision and strategic adaptability are key.

Suplemental Material Statement Datasets used in this paper are still confidential as is being used in R&D activities. Same data model of marketing data sets may be found in ZENODO repository [7]. Source code for SEMTUI and ONTOREFINE are available from EMD Github [8]. Finally, InMa categorization models are considered as confidential as they are being used in commercial marketing optimization actions.

[7] https://zenodo.org/records/3693659
[8] https://enrichmydata.github.io/toolbox/

Acknowledgments. This work has been co-funded by enRichMyData project, funded by European Union's Horizon Europe research and innovation programme, grant No. 101070284, and InMa, funded by Spanish Ministry of Econmic Affairs and Digital Transformation, reference 2021/C005/00141700.

References

1. Artun, O., Levin, D. Predictive marketing: Easy ways every marketer can use customer analytics and big data. John Wiley & Sons (2015)
2. Breslin, J. G., Passant, A., & Decker, S. The social semantic web. Springer Science & Business Media. (2009)
3. Gryllakis, F., Pelekis, N., Doulkeridis, C., Sideridis, S., Theodoridis, Y. (2017). Searching for spatio-temporal-keyword patterns in semantic trajectories. In Advances in Intelligent Data Analysis XVI: 16th International Symposium, IDA 2017. Proceedings 16 (pp. 112-124), Springer International Publishing, London, UK (2017)
4. Radas, S., Shugan, S. M. Seasonal marketing and timing new product introductions. Journal of Marketing Research, **35**(3), 296-315 (1998)
5. Zhani, M. F., Elbiaze, H., Kamoun, F. (2009). Analysis and Prediction of Real Network Traffic. J. Networks, **4**(9), 855-865 (2009)
6. Cutrona, V., De Paoli, F., Košmerlj, A., Nikolov, N., Palmonari, M., Perales, F., Roman, D. Semantically-enabled optimization of digital marketing campaigns. In The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Proceedings, Part II 18 pp. 345-362. Springer International Publishing. Auckland, New Zealand (2019)
7. Chen, A., Koehler, J. R., Owen, A. B., Remy, N., Shi, M. Data enrichment for incremental reach estimation. Technical report, Google Inc. (2014).
8. Ripamonti, M., De Paoli, F., Palmonari, M. Semtui: a framework for the interactive semantic enrichment of tabular data (2022). URL https://arxiv.org/abs/2203.09521
9. Avogadro, R., Ciavotta, M., De Paoli, F., Palmonari, M., Roman, D., Estimating Link Confidence for Human-in-the-loop Table Annotation, in: IEEE/WIC International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Venice, Italy, (2023)
10. Krasteva, I., Petrova-Antonova, D., De Paoli, F., Hristov, E., Borukova, M., Ciavotta, M., Avogadro, R. Geospatial Enrichment of Urban Data for Advanced City Planning: a Pilot Study. In Proceedings - 2023 IEEE International Conference on Big Data, BigData 2023 (pp.3139-3143). IEEE (2024)
11. Zheng, B., Yuan, N. J., Zheng, K., Xie, X., Sadiq, S., Zhou, X. Approximate keyword search in semantic trajectory database. In 2015 IEEE 31st International conference on data engineering pp. 975-986. IEEE.(2015)
12. Larionov, D., Shelmanov, A., Chistova, E., Smirnov, I. (2019, September). Semantic role labeling with pretrained language models for known and unknown predicates. In Proceedings of the International Conference on Recent Advances in Natural Language Processing pp. 619-628 (2019)
13. Tirumala, S. S., Narayanan, A. Hierarchical data classification using deep neural networks. In Neural Information Processing: 22nd International Conference, ICONIP 2015, Istanbul, Turkey, November 9-12, 2015, Proceedings, Part I 22 pp. 492-500. Springer International Publishing. (2015)