

# Automatic detection of potential customers by opinion mining and intelligent agents

Raúl Moreno  
Madox Viajes  
Rey Juan Carlos University  
Arroyomolinos (Madrid), Spain  
r.morenoi@alumnos.urjc.es

Alberto Fernández-Isabel  
Rey Juan Carlos University  
Data Science Laboratory  
Móstoles (Madrid), Spain  
alberto.fernandez.isabel@urjc.es

Isaac Martín de Diego  
Rey Juan Carlos University  
Data Science Laboratory  
Móstoles (Madrid), Spain  
isaac.martin@urjc.es

Javier M. Moguerza  
Rey Juan Carlos University  
Data Science Laboratory  
Móstoles (Madrid), Spain  
javier.moguerza@urjc.es

Carmen Lancho  
Rey Juan Carlos University  
Data Science Laboratory  
Móstoles (Madrid), Spain  
carmen.lancho@urjc.es

Marina Cuesta  
Rey Juan Carlos University  
Data Science Laboratory  
Móstoles (Madrid), Spain  
marina.cuesta@urjc.es

**Abstract**—Customer acquisition is an issue that continues to receive attention from companies worldwide. Various marketing campaigns using psychological methodologies have been designed to address this issue. However, once a campaign is launched, it is highly complicated to detect which sets of customers are most likely to purchase an offered product. This fact is key since it allows companies to focus their efforts on specific clients and discard others. Several selection techniques have been implemented, but most of them are usually very demanding in terms of time and human resources for the companies. Artificial Intelligence techniques appear to help to simplify the process. Thus, companies have started to use Machine Learning (ML) models trained to efficiently detect those clients with certain proneness to purchase. Toward this goal, this paper presents a novel purchase propensity detection ML system based on Sentiment Analysis techniques able to consider customer comments regarding the offered products. The tourist domain was selected for the case study, where the obtained product was successfully embedded in an initial prototype.

## I. INTRODUCTION

CUSTOMER acquisition is one of the most important tasks companies undertake to promote their products and services. It allows expanding the business, enhancing their potential in the markets and producing benefits over time. However, companies need a clear a solid plan of action to be successful in custom acquisition. Creating such a plan in turn requires first-hand knowledge of customers' needs and tastes [1].

Several studies have been conducted to detect these customers' needs, producing behavioral profiles and other different psychological approaches. These proposals work acceptably when companies have specific employees or departments dedicated to making evaluations of the customers. In the case of small or medium enterprises with few resources, or those with high amounts of customers to evaluate, this task is difficult to achieve, becoming a very demanding process.

This work was supported by Madox Viajes Travel Agency

On the other hand, the Internet makes it easy to access collected information about customers' opinions on purchases. Thus, these data can be analyzed to detect patterns that indicate a certain propensity to buy some specific products offered by a company [2]. Specifically, customers whose search is more alternative-based are found to have a higher propensity to purchase than customers whose search is more attribute-based. The main rationale is that customers using alternative based search are evaluating products one at a time, and they are more able to judge whether the product meets their purchasing criteria or it is suitable for them with lesser distraction from other products or services.

In the case of the tourism domain, it is very difficult to build client loyalty due to the wide amount of offers and the fluctuations in the market. Moreover, the recent COVID-19 outbreak has provoked more complications for tourist companies due to the restrictions and measures imposed by governments of countries all over the world. As a result, the acquisition of customers (whose numbers reduced drastically during the period) and the detection of those who are prone to purchase has become a fundamental issue, especially for small or medium-sized companies. Due to this situation, this process is also more demanding and time-consuming, so it is important to develop specific software to support the decision-making step and provide recommendations automatically.

This paper presents a Machine Learning (ML) module able to automatize the detection of potential customers. It has been successfully embedded into an initial prototype of a tourist management framework. The fusion of both elements (i.e., the tourist domain and ML) has produced a novel and complete system based on expert knowledge to manage tourist events and provide recommendations from the perspective of the customer and the travel agency. This enhancement allows the tour companies to reduce efforts and focus their resources on specific individuals and tourist offers.

Regarding the ML module, it analyzes the opinions in texts written by potential customers for different tourist offers, being able to detect the real interest of the individuals. This allows discarding those customers with less predicted interest. Intelligent agents following a Multi-Agent System (MAS) architecture have been included to establish communication and knowledge interchange. This MAS promotes the distribution of the workload and eases the evaluation of multiple texts. This issue is relevant because companies usually need real-time analysis and responses to satisfy the requests made by customers.

Several experiments in real environments have been implemented to show the viability of the proposal. *Madox Viajes* is a tour company that has put in production the developed software, including the ML module with very successful and promising results.

The rest of the paper is organized as follows. Section II situates the approach in the domain and makes some comparisons with similar ideas. Section III details the architectures of the ML module and the prototype. Section IV illustrates a battery of experiments achieved in a real environment. Finally, Section V concludes by making a detailed analysis and provides future guidelines.

## II. BACKGROUND

This section introduces the foundations of the ML module specifically designed to evaluate the opinions of customers. Opinion Mining techniques are put into the spotlight for the case of evaluating opinions. Then, intelligent agents and MAS are detailed focusing on the distribution of the workload. Finally, a first analysis addresses the problematic in tourism related to the detection of prone to purchase customers (see Section II-C).

### A. Sentiment Analysis for the evaluating opinions

Sentiment Analysis (also called Opinion Mining) is one of the most relevant areas in the Natural Language Processing (NLP) domain. Its main objective is to gather the subjectivity expressed by humans in texts [3]. Thus, it measures the influence of a written text over the reader regarding the type of feelings risen and the level of affection.

There are two main perspectives in NLP to address the Sentiment Analysis issue: dictionary-based approaches and ML approaches.

Dictionary-based approaches (also called lexicons) consist of a collection of predefined words which usually appear associated with a sentiment score or a polarity (positive, negative or neutral). They are used to calculate the sentiment polarity of a sentence detecting words included in the dictionary and averaging the polarity of these words. It can be also extrapolated to paragraph or complete texts. However, it has to be considered the noise generated by the set of relevant words (e.g., substantives, verbs or adjectives) that are not detected, and the changes in the discourse made by the creator of the textual content. Well-known examples of these dictionaries are SenticNet [4] and SentiWordNet [5].

ML approaches predict the sentiment values of the words by using statistical models based on distributional semantics (e.g., word embeddings and transformers). These models are built through a training phase on which a collection of words with their corresponding polarity is used. This collection is a corpus that contains the ground truth (i.e., the reality that ML models must try to simulate adapting their parameters to learn). Subsequently, two more steps are usually included: the validation and the test phases. Both allow measuring the quality of the model regarding the ground truth (i.e., the learning capability).

Delving into the ML solutions, there exist several types of approaches to address the Sentiment Analysis task [6]. For instance, there are simple solutions that only use a basic ML discarding the processing step [7]. Other approaches use NLP techniques at the beginning of the pipeline and different embedded ML models later to produce a more complex model [8]. In the first ones, Deep Learning approaches are usually the most typical ones. They do not consider rendering the textual context as they are focused on detecting the syntax and semantics patterns of the text. Common instances of these models are word embedding-based approaches [9] as Word2Vec and LSTM neural networks. However, their performance cannot be compared to the quality provided by the most recent models implemented by bidirectional transformers and attention methods (e.g., BERT [10] and their related approaches). These ML models usually include two attention methods to estimate the value of their parameters: intra-attention and global attention [11]. The first estimates the similarity between words in a sentence, while the second follows a global perspective taking into account the whole textual content.

In the case of the Sentiment Analysis focused on opinions and reviews, the approaches are usually ML solutions adapted to the specific context. However, these proposals are usually limited to only make an estimation and later their results are complemented by the knowledge of human experts. Typical instances of this perspective are: course evaluations [12], online purchase evaluations [13] and movie reviews [14].

The proposed ML module achieves the Sentiment Analysis task, and later gathering conclusions from the results through the distributed intelligence provided by the intelligent agents. Thus, the tourism domain prototype where it is embedded is able to predict which individuals are prone to purchase tourist offers and events.

### B. Intelligent agents to distribute the workload

Intelligent agents are software abstractions able to simulate interactions (with an environment or with other agents) and behaviors from the real world. These elements are proactive, autonomous, and independent, addressing different problems according to the set of predefined rules and knowledge. Their ability to interact can be exploited to solve complex problems or to distribute the workload with certain coordination. Thus, MAS appear as a possible solution. These MAS are usually designed following a level of abstraction. Agent-Based Modeling (ABM) [15] and Agent-Oriented Software Engineering













Concept	Meaning	Icon
Agent	An active element with explicit goals that is able to initiate some actions involving other elements of the simulation.	
Role	A specific collection of tasks performed by agents when pursuing a goal or offering some service to the other members of the society. It has as a result a specific behavior.	
Environment Application	An element of the environment. Agents can act on the environment using its actions and perceive information through its events.	
Goal	An objective of a role/agent. Roles/agents try to satisfy their goals executing tasks. A goal is achieved or fails if some elements (i.e. frame facts and events) are present or absent in the agent groups or the environment.	
Task	A capability of a role/agent. To execute a task, certain elements (i.e. frame facts and events) must be available. The execution produces/consumes some elements as result.	
Frame Fact	An element produced by a task, and therefore by the roles/agents.	
Mental State	Part of the internal state of a role/agent. It groups goals, frame facts and events, and specify conditions on them.	
Belief	Part of the mental state. It contains the knowledge (specific, role-based or general) that an agent possesses. This knowledge can be used to interact with its surrounded environment.	
Conversation	Communication between two or more agents to exchange information. One of these agents plays the role of initiator of the conversation, while the others are the receptors.	
Plan	Representation of the means by which a goal can be satisfied. It is usually structured to provide a deterministic meaning to the operations.	
Mental State Manager	Part of the internal state of a role/agent. It provides for operations to create, destroy and modify mental entities.	
Mental State Processor	Part of the internal state of a role/agent. It determines how a mental state evolves, described in terms of rules or planning.	

Fig. 1. Main concepts of INGENIAS for developing MAS.

(AOSE) [16] provide the elements and entities to tackle this issue.

Intelligent agents present a life-cycle focused on satisfying a collection of goals through several associated concepts (see Fig. 1). These goals are structured in a hierarchical way, having sets of sub-goals that accomplish other goals at higher levels. Goals are associated with a set of tasks that are executed by the agents. Both goals and tasks are part of the mental state of the agents. This mental state plays the role of the brain, containing rules and specific knowledge from the environment that support the execution of the tasks and the satisfaction of the goals.

Agents can take advantage of their ability to interact with the environment to solve complex problems having partial or reduced knowledge about a problem. The organization in MAS opens the collaboration, the competition and also the negotiation. Well-known approaches that use MAS to solve complex problems or simulate real environments are road traffic simulations [17], distributed decision support systems [18], bio-inspired ML-based systems [19], and computer games

[20].

MAS are usually modeled using specific artifacts and entities to tackle the development steps of complex systems. It allows generating graphically relationships and interactions between agents and later transforming them to source code automatically. INGENIAS, GAIA, Prometheus, and Tropos are well-known agent modeling methodologies in charge of providing support through specific languages [21].

Regarding the implementation of MAS, agent platforms are the typical solution. These proposals include features to ease the distribution of the agents and manage their communication channels. Standing out approaches in this area are JADE [22] and MESA [23].

In the presented approach, the selected agent methodology has been INGENIAS, adapting the agent model to the MESA framework. It uses a bio-inspired distribution model based on the behavior of ant colonies that produce a MAS organized in several cumuli of agents (anthills) working together to achieve the Sentiment Analysis task and the evaluation of customers to detect who are prone to purchase.

### C. Customers in tourism

Tourism is one of the most important economic activities all over the world. Everyday, million people are traveling to different destinations in a wide bunch of means of transport: airplanes, trains, cars, etc. This movement of people is usually related to work activities and tourist events. In the first case, individuals manage their travels by themselves (e.g., commuters), while in the second, it is usually addressed by tour companies (e.g., holidays and honeymoons). These companies provide counseling and support to the customers, being responsible of managing and coordinating the different steps during the travel and their stay [24].

Tourism is a volatile active, where several changes in the prices and in the market affect the benefits of the companies, as well as the configurations of travels and its features. World events like the COVID-19 outbreak or the Ukrainian war are recent situations that have produced hard modifications in the tourist sector. For this reason, the identification and capture of new customers, and the detection of those ones that are prone to purchase is basic for companies to optimize their workload.

Delving into the proneness of customers to purchase a tourist activity, a wide range of variables motivates and make some particular purchase decisions [25]. Instances of the most typical variables are: the particular culture, the emotional and physical state, and personal issues (e.g., visit a friend or a relate). These variables are called as motivators and can be classified into six main categories: primary motives, secondary motives, rational motives, emotional motives, conscious motives, and dormant motives [26].

Primary motives are those situations that force a person to purchase a tourist offer (e.g., a health problem with a relative who is in another country), and secondary motives consists of situations that modify the choice of the customers (e.g., the price of a flight according to different companies). In the case of the rational and emotional motives, the first is

related to an objective evaluation of the customer (e.g., a group of several members rents a minibus instead of some particular vehicles), while the second is the opposite, the customers are moved by their feelings (e.g., a person is prone to purchase a flight with a company instead of another one due to only personal preferences). Conscious motives encompass those ones where customers are aware of their personal needs (e.g., customers have to travel to a place where there is not a train station, therefore renting a car could be a better option). Finally, dormant motives are those ones that are unconscious and usually related to the influence of the society over the customers (e.g., customers have to travel to a place where individuals living there usually have high incomes, then the customers prefer to rent a car instead of traveling by bus).

To address and detect these motivations, the knowledge of psychological experts and the experience provided by travel agents become very relevant. However, when a company has to evaluate hundreds of proposals each day, the human evaluation is almost impossible. Thus, different systems and models to mitigate the problem has been developed. Some of them are very specific in the tourist domain and consider only behavioral features [27]. Other approaches are focused on the impact of the current technologies in tourism and how to adapt the offers to the high connectivity world nowadays [28]. And finally, there are others that evaluate the obtained results of the companies avoiding to consider the opinion and the level of satisfaction of customers [29].

In the case of the presented approach, customers are put into the spotlight, evaluating their opinions before their purchase, knowing their proneness and preferences for the different offers proposed. It allows developing a recommendation process for customers, and it eases the work of the travel agencies being able to discard and focus on specific customers in a smart way.

### III. PROPOSED FRAMEWORK

This section details the novel ML module specially designed to evaluate the opinions of potential customers regarding a set of possible tourist offers, and the initial prototype of the tourist framework where it is embedded.

The ML module allows companies to reduce the effort in time and human resources, as employees can focus only on the customers that are more prone to purchase a product. This functionality works by analyzing the textual comments provided by the customers through Sentiment Analysis techniques and MAS to distribute the workload of the system.

In the case of the tourism domain, the prototype consists of a system focused on the two main perspectives related to the tourism market: customers and travel agents. It provides some features to manage the roles independently covering all the processes from the customers' perspective and the travel agents' perspective. Moreover, it can organize the different interactions between users with both roles (e.g., a tourist offer managed by a travel agent and a customer both using the system). From the customers (i.e., the tourists), the system can make recommendations according to specific preferences

through similarity comparisons based on their profile, as well as modifications through filtering processes. On the other hand, from the travel agents, the system includes the tourist services and offers available during the day, the ratings, the feedback of former customers about the services, and a novel functionality to indicate the most prone to purchase customers.

Next section provides further details about the architecture of the ML module. Next, the design of the prototype is tackle focusing on the different modules of the system and highlighting how the ML module is included in the release.

#### A. Propensity to purchase detection ML module

The ML module is organized into two main components: the *module administrator* and the *Sentiment Analysis administrator*. Both components work together internally (see Fig. 2).

The first component is the manager of the complete module, being in charge of obtaining the stored opinions of customers, managing the *Sentiment Analysis administrator* component, and producing the final results from the outcome of the former component. It comprehends three different entities: the *texts gatherer*, the *workload manager* and the *result processor*.

The *texts gatherer* collects the opinion to be processed by the module. They are loaded from the *Services information* database. This element cleans the text and applies corrections to misspelled words.

The *workload manager* organizes the texts using a Kafka queue and checks if the independent anthills are busy. It can create new MAS on-demand automatically or implement load balancing politics. The configuration of the ML module consists of 4 anthills by default.

The *result processor* obtains the outcomes from the anthills and organizes them to be stored in the database. Thus, the opinion of the customers regarding a tourist offer or service is labeled as positive or negative. Then, the system can select those customers prone to purchase.

The second component is the container of the anthills. These latter are bio-inspired hierarchical structures organized to distribute the workload between several agents. Therefore, the system is ready for processing the information provided by a big amount of potential customers. These anthills are MAS formed by a queen and sets of soldiers and workers, organized in a similar way to a real ant family structure. The number of agents playing the roles of the soldiers and workers can be modified by the system, though they are prefixed to 5 and 10 respectively. The queen agent is fixed by default to one per anthill, and that fact cannot be modified in the present release of the framework. The component is completed with the previously trained ML model that encompasses techniques to achieve the Sentiment Analysis task.

Delving into the Sentiment Analysis model, initially, the Universal Sentence Encoder (USE) was used for generating embeddings from the customer comments. Then, a Convolutional Neural Network (CNN) with the next sequential model pipeline is used: an input 256 dense layer by using ReLU activation, a dropout layer with a rate of 0.5, a 128 dense

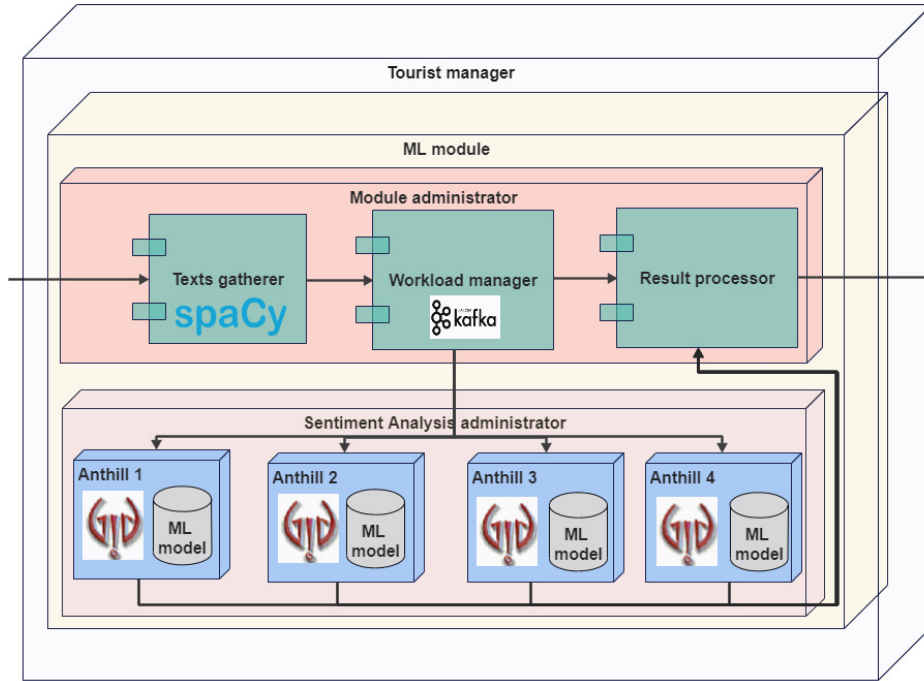


Fig. 2. Architecture of the ML module inside the *tourist manager* with the default configuration.

TABLE I  
SUMMARY OF THE CNN CONFIGURATION.

Layer pipeline	Configuration
ReLU dense layer	256
Dropout	0.5
ReLU dense layer	128
Dropout	0.5
Softmax dense layer	2

layer by using also ReLU activation, another dropout layer with rate 0.5 and the end layer with a 2 dense layer by using a Softmax activation (see Table I).

Regarding the intelligent agents, they are the independent entities that process the texts through cooperative activities. A MAS consists of an anthill that encompasses three types of roles for the agents: queen, soldiers, and workers (see Fig. 3). Therefore, it is organized following bio-inspired ant social structure. The queen (i.e., the manager agent) is responsible for the anthill, receiving the texts to analyze directly from the *workload manager*. Notice that several texts can be sent to this agent. Then, the queen agent assigns the activity of evaluating the text to one soldier agent (i.e., the evaluator agent). This agent is in charge of assigning pieces of text to promote the distribution (usually paragraphs) or complete texts to the worker agents. These latter process the text through the Sentiment Analysis model provided to the anthill (each anthill has a copy of the ML model). Once worker agents have concluded their task, the soldiers join the texts if necessary (protecting and supervising the result obtained by a worker or a set of them) and they return the result to the queen agent.

This process follows the Belief-Desire-Intention (BDI) model [30] where each agent present a goal or a set of them to satisfy to complete its life-cycle. In this sense, the queen has the *assignedtask* goal, the soldiers have the *evaluatedopinion* goal and the workers have the *processedtext* goal. All the goals usually include associated tasks that are actions to achieve. Notice that the agents present at least one task that solves their corresponding goals. These tasks are applied in the environment that agents share. In this case, the environment is the current opinion or set of them of the possible customers. Each agent incorporates a mental state and a set of beliefs (motivations). The queen has in them simple rules to manage the texts, while soldiers present similar rules to distribute the text between the workers. However, workers include the ML model in their mental states to achieve the evaluation of the texts and some simple rules to organize the process in the beliefs. Finally, interactions between the individuals follow the hierarchical structure. They are completed through direct conversations. Notice that in this case, workers do not need to establish conversations with other workers since they tackle their commitment individually according to the orders of the soldier.

The design of the anthill model has been addressed through the INGENIAS agent methodology. Then, the resulting composition has been transformed to be compliant with the MESA framework. The conversations and interactions of the agents have been implemented following the Foundation for Intelligent Physical Agents (FIPA) standards [31].

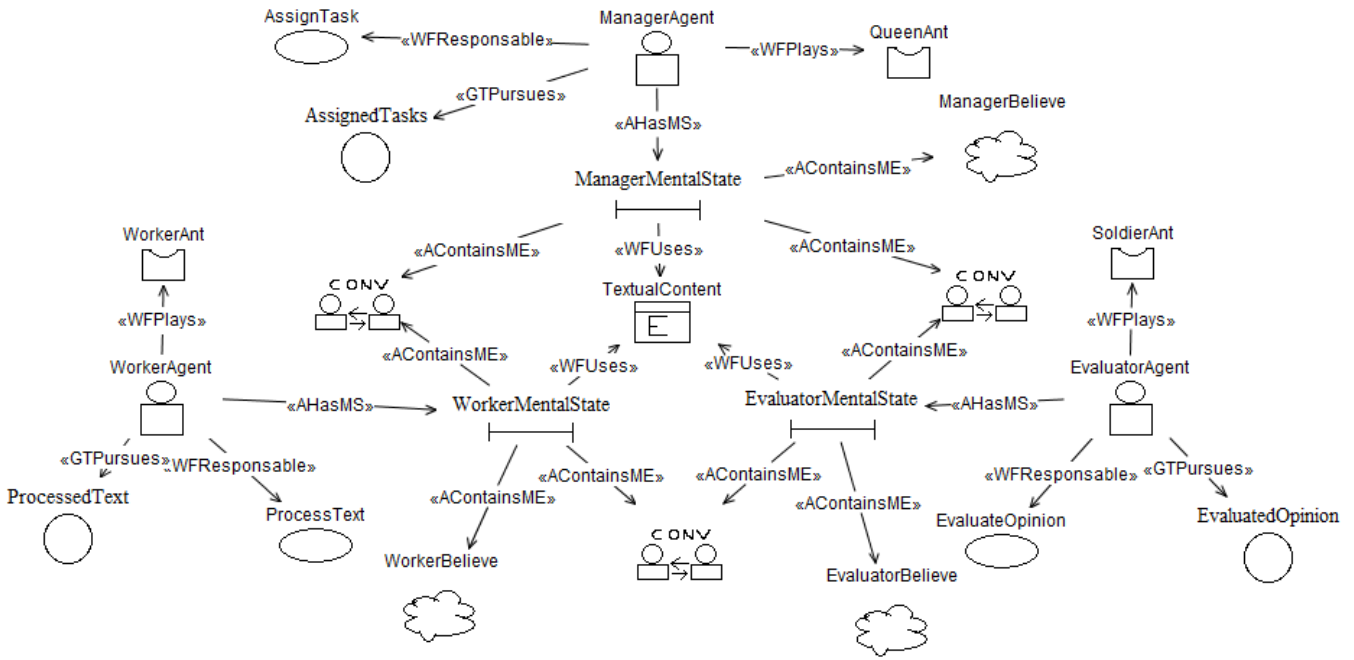


Fig. 3. Excerpt of the main entities involved in the anthill MAS.

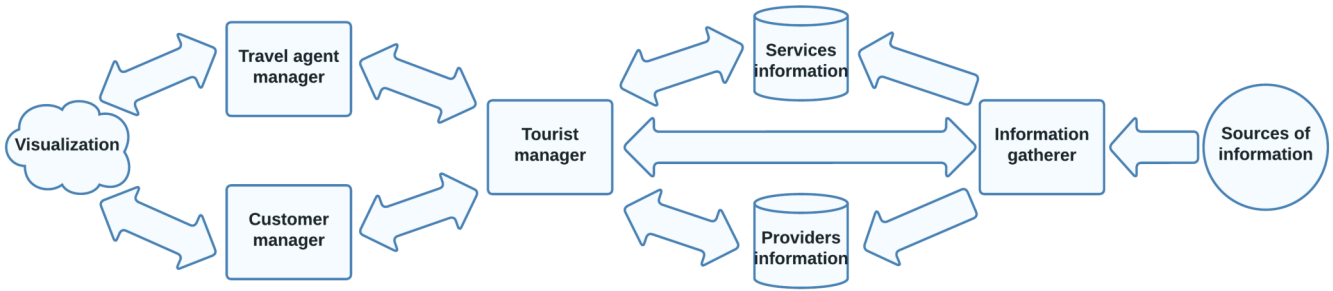


Fig. 4. Overview of the main architecture of the prototype.

**B. Architecture of the prototype**

The architecture of the prototype comprehends two databases and four main modules. These elements are related between them interacting automatically to recover information or responding the requests made by users through the graphical interface (see Fig. 4).

Regarding the databases, they are: *services information* and *providers information*. The first stores the information of the tourist services managed by the company. Both final users (i.e., independent customers and travel agents) produce relevant data that is stored here. It also contains a historic to generate different visual forms for the business. Note that information about customers and their opinions are also placed in that database. Therefore, it is used by the novel ML module to obtain textual content and store the obtained results. On the other hand, the second conserves the knowledge about the providers (i.e., the different companies that daily offer tourist services). That information is updated every day through an

automatic process but it needs the support of an expert in the domain to accomplish some specific and complex tasks related to the several offers and modes. The visual interface of the system provides specific graphical assistants and the architecture presents some modules to properly interact with both databases.

In the case of the main modules, they are: *information gatherer*, *tourist manager*, *travel agent manager* and *customer manager*. Notice that the first two modules are generic to every operation in the system, while the other two are specific for travel agents and customers respectively.

The *tourist manager* module is the core of the framework. It contains information about the opportunities and the profiles and preferences of the final users. With this information, it can produce recommendations to provide the best tourist resources for a trip. It is the module of the system that embeds the proposed ML module. It allows the system to select the most interesting customers according to the evaluation made through the analysis of their opinion.

The *information gatherer* module recapitulates and processes the information from the virtual tourism market. This market reflects the exchanges of information between the providers and tour companies. This module collects static and dynamic information. The first is information that is not frequently updated (e.g., the name of a hotel and its description). The second is volatile information that fluctuates several times during the day (e.g., occupancy levels or rates). Notice that some of this information cannot be stored in the system due to legal issues, so it is always consulted directly from web information sources.

The *customer manager* module is in charge of processing the events and interactions made by the customers. It provides information to the *tourist manager* module to collect information about a specific individual or tourist services and adequate offers depending on the preferences of the customers. Notice that the operations made by individuals with that role must be approved by a travel agent at some point in the pipeline before formalizing a possible trip.

The *travel agent* module provides assistance to travel agents to consult, modify and also create tourist offers. It also supports the selection of specific tourist resources to configure a trip. To achieve that task, this module makes use of the *tourist manager* module. Moreover, the filtering process of prone to purchase customers is also used here to show that knowledge to the different travel agents on demand.

#### IV. EXPERIMENTS

This section details the experiments achieved with the prototype of the framework in a real environment. This environment has been provided by *Madox Viajes*, a tour company that has implanted the current version of the system.

The experiment starts training the ML model used by the system. In this sense, a dataset provided by the company has been used. It consists of a set of 11,840 labeled observations being divided into 90% for training and 10% for testing purposes. These data is unbalanced since the most of opinions are from opportunities that were not bought by customers. Thus, the range of the texts in the dataset is 60% (not bought) and 40% (bought). The decision of using an unbalanced dataset is related to the idea of reducing the false positive predictions as in real environments as the customers are more likely not to purchase.

The output of the ML model is the probability of a tourist opportunity offered by the company of being purchased. Therefore, the company's employees could use this insight when working on opportunities. It should be noted that the focus on employees is important, as it requires really explaining to them what it means to "purchase" a tourist opportunity. In this case, it was necessary to make them understand that it is a probability that can guide them to focus on the best opportunities and discard or defer the rest. Thus, this probability was binarized to simplify this information to the company's employees. Several thresholds for the probability have been tested. The performance results are presented in Table II. In

this case, the optimal threshold was 0.1, corresponding to a  $F_2$ -score equals to 0.8037, Precision 0.8488 and Recall 0.7932, respectively. That is, the ML model is right almost in the 85% of times when the prediction given by the model is "purchase". On the other hand, the ML model is able to recover more than the 79% of all the sold opportunities. Notice that the selection of the  $F_2$ -score is motivated by the fact that for the company it is more relevant higher values of Recall to reduce efforts and increment the benefits.

After deploying the model in production, the ML module were embedded in the system with a default configuration of four anthills of intelligent agents. Then, the team of the company was trained in the use of the scoring produced by the ML model.

The system has been working during several months in the company. It was planned to measure the effect of scoring on the sales pipeline 8 months later (i.e., it was deployed on September 1<sup>st</sup>, 2021). Thus, the results were shown in the yearly harvest report for opportunities on April 1<sup>st</sup>, 2022 (see Table III). It is necessary to highlight that the 2021 sales were aligned with the 2020. It implies a normal behavior, as the company continued operating under pandemic circumstances. In this case, percentages over the total number of opportunities are used to avoid confidential data of the company that could provide insights to competitors. The most relevant information in order to evaluate the performance of the proposed ML module is the percentage of opportunities created in 2021 and sold in 2022: 2,78%. Notice that this is the highest percentage of opportunities sold in one year and were created in the previous year. This means that the scoring is really impacting positively on the travel agents. In fact, Table IV shows the percentages of opportunities created in a year and sold in the same year or in the next year. For instance, given the total amount of opportunities created in 2017 that were sold, the 89.2% were sold in the same year and the remaining ones (9.8%) were sold in 2018. In 2022, the effect of the ML model can be visualized. Thus, almost half of the tourist opportunities created in 2021 have been sold during the first four months of 2022 (41.9% which is almost more than 10 times increase compared to the previous year's value), being this number the highest in the history of the corresponding tourist company.

In conclusion, the incorporation of the proposed ML model in the framework has increased the detection of opportunities of interest for new customers. This fact has been translated into better business statistics, reducing the effort of the company in the creation of new packages, and the reduction of time demand for the employees.

#### V. CONCLUSIONS

This paper has presented a ML module created to evaluate the comments about different tourist offers and services, and to measure the propensity of potential customers to purchase them. It allows classifying the customers according to their proneness of completing a booking (i.e., it detects the most interesting customers). The module has been embedded in a prototype of a framework specifically designed for tourist

TABLE II  
PERFORMANCE OF THE ML MODEL FOR DIFFERENT THRESHOLDS FOR THE PROBABILITY OF PURCHASE.

Metrics / Threshold	0.01	0.1	0.2	0.3	0.4	0.5
<b>Precision</b>	0.5476	0.8488	0.8920	0.9123	0.9299	0.9386
<b>Recall</b>	0.8639	0.7932	0.7660	0.7349	0.7141	0.6945
<b>Accuracy</b>	0.8892	0.9546	0.9574	0.9562	0.9557	0.9543
<b>F<sub>2</sub>-score</b>	0.7744	0.8037	0.7882	0.7646	0.7488	0.7326

TABLE III  
YEARLY HARVEST FOR PREDICTED PURCHASE AND PURCHASE OPPORTUNITIES ON APRIL 1<sup>st</sup>, 2022.

Year Created / Year Sold	2017	2018	2019	2020	2021	2022	Total
<b>2017</b>	14, 35%	1, 70%					16, 05%
<b>2018</b>		30, 94%	0, 72%				31, 66%
<b>2019</b>			35, 34%	0, 81%			36, 14%
<b>2020</b>				8, 43%	0, 36%		8, 79%
<b>2021</b>					3, 86%	2, 78%	6, 64%
<b>2022</b>						0, 72%	0, 72%
<b>Total</b>	14, 35%	32, 65%	36, 05%	9, 24%	4, 22%	3, 50%	100, 00%

TABLE IV  
COMPARISON BETWEEN CREATED AND SOLD OPPORTUNITIES ON APRIL 1<sup>st</sup>, 2022.

	2018	2019	2020	2021	2022
<b>Opportunities created and sold in the previous year</b>	89.2%	97.7%	97.8%	95.6%	58.1%
<b>Opportunities created in the previous year and sold in the current year</b>	9.8%	2.3%	2.2%	4.4%	41.9%

management from both perspectives: the customer and the travel agent.

The ML module analyzes the opinion of the possible customers through Sentiment Analysis techniques based on neural networks. Moreover, it includes a bio-inspired architecture design of MAS to distribute the workload of the system. The complete system has been deployed in a real environment, showing its ability to manage a real tour company. The obtained results have been highly satisfying; the system has increased the ability of the company to find new potential customers. These results have translated into more economical benefits, simplification of the creation of new tourist opportunities, and less time spent by the human resources.

In the future, the MAS organization will be stressed with a test battery to find possible weaknesses of the architecture. Regarding the ML module in general, it will be improved by incorporating other techniques like Weight of Evidence (WoE). This fact will produce a novel and complete release of the system for the tourist domain. The architecture of the framework will be also improved and later adapted to be used through micro-services. All these upgrades will lead to a complete deployment where travel agents and customers will have access in real-time to the provided functionalities.

#### ACKNOWLEDGMENT

This research has been supported by the grants from Madrid Autonomous Community (Ref: IND2019/TIC-17194), the Spanish Ministry of Economy and Competitiveness, under the Retos-Investigación program: MODAS-IN (Ref: RTI-2018-094269-B-I00), and Rey Juan Carlos University (Ref: C1PREDOC2020); and donation of the Titan V GPU by NVIDIA Corporation.

#### REFERENCES

- [1] J. S. Thomas, "A methodology for linking customer acquisition to customer retention," *Journal of marketing research*, vol. 38, no. 2, pp. 262–268, 2001. doi: 10.1509/jmkr.38.2.262.18848
- [2] O. Mintz, I. S. Currim, and I. Jeliakov, "Information processing pattern and propensity to buy: An investigation of online point-of-purchase behavior," *Marketing Science*, vol. 32, no. 5, pp. 716–732, 2013. doi: 10.1287/mksc.2013.0790
- [3] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-based systems*, vol. 89, pp. 14–46, 2015. doi: 10.1016/j.knosys.2015.06.015
- [4] E. Cambria, Y. Li, F. Z. Xing, S. Poria, and K. Kwok, "Sentinet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis," in *Proceedings of the 29th ACM international conference on information & knowledge management*, 2020. doi: 10.1145/3340531.3412003 pp. 105–114.
- [5] N. Medagoda, S. Shanmuganathan, and J. Whalley, "Sentiment lexicon construction using sentiwordnet 3.0," in *2015 11th International Conference on Natural Computation (ICNC)*. IEEE, 2015. doi: 10.1109/ICNC.2015.7378094 pp. 802–807.
- [6] M. Ahmad, S. Aftab, S. S. Muhammad, and S. Ahmad, "Machine learning techniques for sentiment analysis: A review," *Int. J. Multidiscip. Sci. Eng.*, vol. 8, no. 3, p. 27, 2017. doi: 10.18090/samriddhi.v12i02.03
- [7] S. Bhattacharya, D. Sarkar, D. K. Kole, and P. Jana, "Recent trends in recommendation systems and sentiment analysis," *Advanced Data Mining Tools and Methods for Social Computing*, pp. 163–175, 2022. doi: 10.1016/B978-0-32-385708-6.00016-3
- [8] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *Journal of Informetrics*, vol. 3, no. 2, pp. 143–157, 2009. doi: 10.1016/j.joi.2009.01.003
- [9] A. Onan, "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 23, p. e5909, 2021. doi: 10.1002/cpe.5909
- [10] M. Pota, M. Ventura, H. Fujita, and M. Esposito, "Multilingual evaluation of pre-processing for bert-based sentiment analysis of tweets," *Expert Systems with Applications*, vol. 181, p. 115119, 2021. doi: 10.1016/j.eswa.2021.115119
- [11] N. Liu, B. Shen, Z. Zhang, Z. Zhang, and K. Mi, "Attention-based sentiment reasoner for aspect-based sentiment analysis," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, pp. 1–17, 2019. doi: 10.1186/s13673-019-0196-3



- [12] A. Onan, "Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach," *Computer Applications in Engineering Education*, vol. 29, no. 3, pp. 572–589, 2021. doi: 10.1002/cae.22253
- [13] S. Wassan, X. Chen, T. Shen, M. Waqar, and N. Jhanjhi, "Amazon product sentiment analysis using machine learning techniques," *Revista Argentina de Clínica Psicológica*, vol. 30, no. 1, p. 695, 2021. doi: 10.24205/03276716.2020.2065
- [14] P. K. Mallick, P. Dutta, S. Mishra, and M. K. Mishra, "Sentiment analysis and evaluation of movie reviews using classifiers," in *Cognitive Informatics and Soft Computing*. Springer, 2021. doi: 10.1007/978-981-16-1056-1\_5 pp. 53–59.
- [15] M. Uddin, Q. Wang, H. H. Wei, H. L. Chi, and M. Ni, "Building information modeling (bim), system dynamics (sd), and agent-based modeling (abm): Towards an integrated approach," *Ain Shams Engineering Journal*, vol. 12, no. 4, pp. 4261–4274, 2021. doi: 10.1016/j.asej.2021.04.015
- [16] A. Garro, M. Mühlhäuser, A. Tundis, M. Baldoni, C. Baroglio, F. Bergenti, P. Torroni *et al.*, "Intelligent agents: Multi-agent systems," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 315, 2018. doi: 10.1016/B978-0-12-809633-8.20328-2
- [17] A. Fernández-Isabel, R. Fuentes-Fernández, and I. M. de Diego, "Modeling multi-agent systems to simulate sensor-based smart roads," *Simulation Modelling Practice and Theory*, vol. 99, p. 101994, 2020. doi: 10.1016/j.simpat.2019.101994
- [18] C. González-Fernández, J. Cabezas, A. Fernández-Isabel, and I. Martín de Diego, "Combining multi-agent systems and subjective logic to develop decision support systems," in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, 2020. doi: 10.1007/978-3-030-50146-4\_12 pp. 143–157.
- [19] J. Cabezas, A. Fernandez-Isabel, R. R. Fernández, C. González-Fernández, A. Alonso, and I. M. de Diego, "Bio-inspired agent-based architecture for fraud detection," in *Proceedings of the 2020 3rd International Conference on Information Management and Management Science*, 2020. doi: 10.1145/3416028.3416039 pp. 67–71.
- [20] J. Wang, Y. Hong, J. Wang, J. Xu, Y. Tang, Q.-L. Han, and J. Kurths, "Cooperative and competitive multi-agent systems: From optimization to games," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 5, pp. 763–783, 2022. doi: 10.1109/JAS.2022.105506
- [21] P. Siswahyudi, T. A. Kurniawan, and V. Sugiarto, "Agent-oriented methodologies comparison: A literature review," *Advanced Science Letters*, vol. 24, no. 11, pp. 8710–8716, 2018. doi: 10.1166/asl.2018.12331
- [22] F. Bergenti, G. Caire, S. Monica, and A. Poggi, "The first twenty years of agent-based software development with jade," *Autonomous Agents and Multi-Agent Systems*, vol. 34, no. 2, pp. 1–19, 2020. doi: 10.1007/s10458-020-09460-z
- [23] J. Kazil, D. Masad, and A. Crooks, "Utilizing python for agent-based modeling: the mesa framework," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, 2020. doi: 10.1007/978-3-030-61255-9\_30 pp. 308–317.
- [24] B. K. Kler, "Tourism and restoration," in *Philosophical issues in tourism*. Channel View Publications, 2009. doi: 10.21832/9781845410988 pp. 117–134.
- [25] A. Vinerean, "Motivators that intervene in the decision making process in tourism," *Expert journal of marketing*, vol. 2, no. 2, 2014.
- [26] J. Swarbrooke and S. Horner, *Consumer behavior in tourism*. Heinemann, Oxford, 2007.
- [27] S. Amaro and P. Duarte, "An integrative model of consumers' intentions to purchase travel online," *Tourism management*, vol. 46, pp. 64–79, 2015. doi: 10.1016/j.tourman.2014.06.006
- [28] R. P. Falcao, J. B. Ferreira, and M. Carrazedo Marques da Costa Filho, "The influence of ubiquitous connectivity, trust, personality and generational effects on mobile tourism purchases," *Information Technology & Tourism*, vol. 21, no. 4, pp. 483–514, 2019. doi: 10.1007/s40558-019-00154-1
- [29] A. A. Mahrour and S. S. Hassan, "Achieving superior customer experience: An investigation of multichannel choices in the travel and tourism industry of an emerging market," *Journal of Travel Research*, vol. 56, no. 8, pp. 1049–1064, 2017. doi: 10.1177/0047287516677166
- [30] L. De Silva, F. R. Meneguzzi, and B. Logan, "Bdi agent architectures: A survey," in *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), 2020, Japão.*, 2020. doi: 10.17863/CAM.53101
- [31] L. S. Melo, R. F. Sampaio, R. P. S. Leão, G. C. Barroso, and J. R. Bezerra, "Python-based multi-agent platform for application on power grids," *International transactions on electrical energy systems*, vol. 29, no. 6, p. e12012, 2019. doi: 10.1002/2050-7038.12012