Zapata, G., Murga, J., Raymundo, C. et al. Business information architecture for successful project implementation based on sentiment analysis in the tourist sector. *J Intell Inf Syst* **53**, 563–585 (2019).

DOI: 10.1007/s10844-019-00564-x

# Business Information Architecture for successful project implementation based on Sentiment Analysis in the tourist sector

**Gianpierre Zapata · Javier Murga · Carlos Raymundo · Francisco Dominguez · Javier M.Moguerza · Jose Maria Alvarez**

Received: date / Accepted: date

**Abstract** In the today's market, there is a wide range of failed IT projects in specialized small and medium-sized companies because of poor control in the gap between the business and its vision. In other words, acquired goods are not being sold, a scenario which is very common in tourism retail companies. These companies buy a number of travel packages from big companies and due to lack of demand for these packages, they expire, becoming an expense, rather than an investment. To solve this problem, we propose to detect the problems that limit a company by re-engineering the processes, enabling the implementation of a business architecture based on sentimental analysis, allowing small and medium-sized tourism enterprises (SMEs) to make better decisions and analyze the information that most possess, without knowing how to exploit it. In addition, a case study was applied using a real company, comparing data

G. Zapata
Escuela de ingeniería de Sistemas y Computación. Universidad Peruana de Ciencias Aplicadas, Lima, Perú E-mail: u201214895@upc.edu.pe

J. Murga
Escuela de ingeniería de Sistemas y Computación. Universidad Peruana de Ciencias Aplicadas, Lima, Perú E-mail: u201111811@upc.edu.pe

C. Raymundo
Escuela de ingeniería de Sistemas y Computación. Universidad Peruana de Ciencias Aplicadas, Lima, Perú E-mail: Carlos.raymundo@upc.edu.pe

F. Dominguez
Escuela Superior de Ingeniería Informática, Universidad Rey Juan Carlos, Mostoles, Madrid, Spain E-mail: francisco.dominguez@urjc.es

J. Moguerza
Escuela Superior de Ingeniería Informática, Universidad Rey Juan Carlos, Mostoles, Madrid, Spain E-mail: javier.moguerza@urjc.es

J. Alvarez
Department of Computer Science and Engineering, Universidad Carlos III, Madrid, Spain E-mail: joalvare@inf.uc3m.es

before and after using the proposed model in order to validate feasibility of the applied model.

**Keywords** Data Governance · Enterprise Architecture · Business Model · Process Improvement · Tourism management.

## 1 Introduction

At present, SMEs in Peru have an increasing presence in the Peruvian market, which generates high demand and operational competitiveness, and cooperative work between all companies in the different areas, therefore each seeks ways to improve their services, reduce costs, and time to improve their internal processes, in addition to finding new technological solutions that may bring great benefits to the organization. On the other hand, with the innovative technologies present in the market, all of which are accompanied by implementation models for a business, meaning we can create new scenarios tailored to the measures and needs of the users. Considering that the majority of SMEs do not have a high profitability, there are tools in the Cloud that mitigate these costs with recovery in the short or medium term, in Peru only 15% of SMEs use technological tools in their business processes [1].

The purpose of this study is to provide an architecture of principle, a strategy that meets the needs of a tourism company in the Peruvian market, so that when the problem arises in the travel management processes, good practices are presented to improve those processes and develop technological solutions. These solutions use Sentiment Analysis in conjunction with the use of large amounts of data which allow us to analyze and predict the trends of users allowing the company to reduce the travel time, to make reservations, and to generate new travel plans, which in itself carries a series of guidelines and plans for a company in this sector to operate adequately and continuously improve tourist processes, thus improving company sales.

In this sense, a new business model is implemented following the guidelines of a business architecture, taking into account that the research will be divided into 5 sections that are described as follows; section 2 presents the state of the art about the concepts theories that support research, section 3 presents the proposed model and its description, section 4 presents what is an architecture, its elaboration, its steps, models and the solution proposed for this problem, section 5 presents the validation in a real case to finish with the conclusions of the execution.

## 2 Background

### 2.1 Information Analysis Model Based on Sentiments

The inception of information analysis models occurs when users interact within networks. In this way, Sentiment Analysis is recognized as the use of text

processing and analysis tools which allow us to quantify the feelings that users are expressing. This helps to discover trends that express real world events in a virtual environment [2], it is also useful for the prediction of different topics [3]. In our case, it is used to predict tourist traffic and the time series of social networks. This allows the generation of new research cases to then propose a predictive analysis for tweets related to tourism and thus be able to complete the parameters in the qualification and location rubric [4].

### 2.2 Predictive Analytics

Predictive Analytics is a set of business intelligence technologies that discover relationships and patterns in large volumes of data that can be used to predict behavior and events. Unlike other business intelligence technologies, predictive analytics offers a view into the future, using past events to predict it [5]. Predictive analytics includes statistical models and other empirical methods for the creation of empirical predictions (as opposed to predictions based exclusively on theory), as well as methods that identify the quality of predictions in practice [6]. It is important to recognize the limitations of predictive analytics. First, progress cannot be made without a data set of adequate size and quality that serves as training. Second, it is crucial to have a clear definition of the concept that is to be predicted and to have historical examples [7]. In this way, the role that a predictive model plays is to collect the available data, classify them according to the defined parameters and after analyzing them, propose an accurate prediction about a future result. To enhance these steps, the parameterization of the data is of utmost importance, since the prediction will depend on these parameters. So, we must define a number of variables that fit the model, since a lack or excess of these will produce biased or less accurate predictions respectively. These variables depend on the chosen environment, so in the case of this study they will be related to tourism [8].

### 2.3 Technological Impact in the Tourism Sector

Both cultural and technological development have encouraged the integration of tourism and technology models [9]. Evidence of this is found in the large number of web resources that promote tourism [10]. With this large amount of information, data mining takes on the important role of converting seemingly useless data into a source of information that indicates user interests and trends [11]. For instance, we can perform analysis of cities to map the behavior of the inhabitants and know how landmarks are prioritized [12]. These technological advances make it possible for cities to be prepared [13] to accommodate constructions and improvements that are both visually and functionally more pleasing to citizens, giving them more opportunities to appreciate the urban landscape [14].
These models can be implemented within a framework that jointly uses the

benefits of semantic networks and cloud services[15], thus it would be possible to extract, classify and analyze user patterns and relationships [16]. Most of the proposals are aimed at e-commerce for the improvement of web portals, but none of them tries to solve the root of the problem. Hence,so we propose a travel management process model for SMEs in the tourism sector in Peru. This tool would present an advantage to companies that do not have a defined process or wish to know the trends and interests of their clients.

## 3 Proposed Model

Based on a literature review, we propose a business model to successful project implementation based on Sentiment Analysis of a case study that presents scenarios where the internal processes of the tourism sector can be used as information sources that allow us to analyze the characteristics and consumer behaviors [17]. It is noteworthy that the model proposed for the case study falls within the five phases of project development, which recognizes a company's needs to carry out operations without having to worry about issues related to the SME's budget. See Fig. 1.

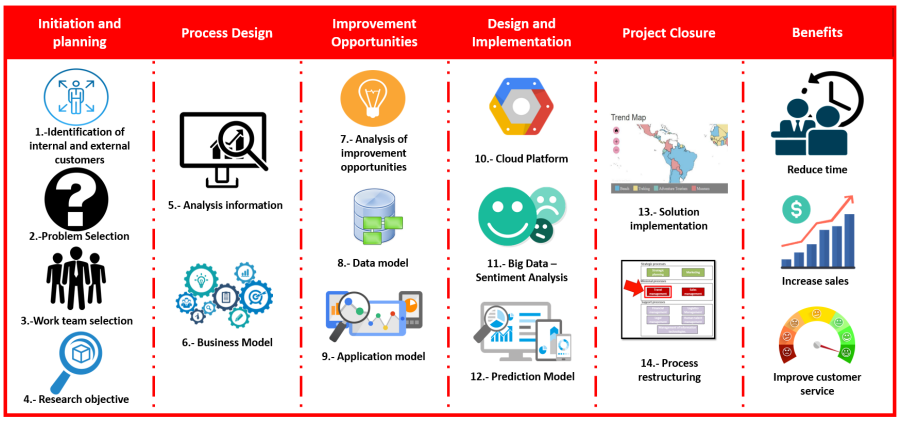


**Fig. 1** Phases to define a business model

Therefore, the model was designed taking the following steps into account: the planning and initiation of the scenario for the execution of the case study, in addition to studying customer behavior to identify the problems affecting the company, identify business processes and describe the relationship between processes, optimize and reorganize the processes, and develop the architectures, ending with the technological proposals. This gives us broad knowledge of the business and implement good practices to define the version of the company's processes that helps improve travel management, thus generating competitive advantages at a commercial and technological level in the context of globalization, competitiveness, and greater participation in the market.

## 4 Conceptual Model

### 4.1 Phase 1: Initiation and planning

#### *4.1.1 Initiation*

**Identification of internal and external customers**
The first step in the process of this case study is to identify which clients the process impacts, so that the internal customer is defined as a person who is part of the process within the organization or institution. For example in this case, a sales persons in marketing, since this person is in charge of spreading the various tour packages that the organization publishes to the market is considered an internal customer. While an external client is anyone that provides a service or product paid for by him.
**Problem Selection**
A problem in the tourism sector can be due to any of the following factors: diversity of travel packages, inadequate logistics planning, level of service, response to out-of-date requests and cancellation of package reservations. The problem is selected according to the criteria defined in the company's policies (Security, Quality, Delivery, Cost or Service level), regarding the work group, the managers, and the results of the same.
**Impact on the company**
At this point we state how process improvement impacts the business. The consequences of not implementing the new business model are mentioned. We must know what the business situation has been due to the current process, and answer the questions presented by the company. What has caused the loss of customers? Why are package reservations being canceled? Why are more packages being reserved than current demand? It is important to describe how the research is aligned with the initiatives and goals of the tourism business. For this reason, in Figure 2. (the Evaluation matrix), we can analyze according to previously defined evaluation criteria and the detected problems that affect SMEs, which is the order of priority to be able to define the problem to be addressed in our case study.
**Problem description**
The company's current situation is analyzed and the question of "Why?" is asked, until we are able to define the problem in the process, product or service specification, indicated quantitatively, demonstrating the need to modify the current state. In this way, resources are put to better use, aiming not to name a problem, but rather understand the cause of the problem and form questions about a definitive solution that meets the research objectives.
**Scope of the investigation**
It is useful to delimit the process.

- Start point: Identify the activity where the process begins.
- End point: Identify the activity where the process ends.
- Within scope: Activities that are within the process.
- Out of reach: Activities that are not within the process

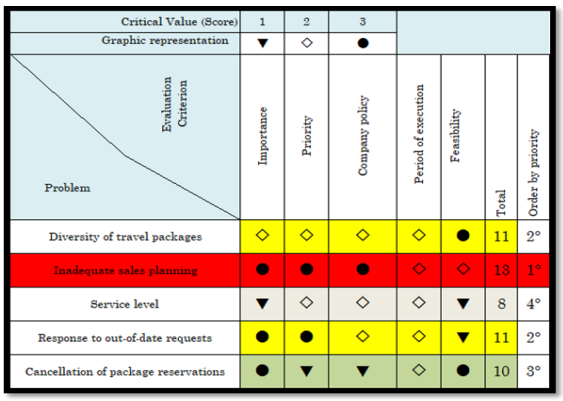


**Fig. 2** Evaluation matrix

In this way, the critical points and activities that must be carried out are defined in order to define the research objectives. See Table 1.

**Table 1** Definition of research objectives

| **Critical Point** | **Activities** |
|---|---|
| Clarify the goal value. Do not simply express the wishes and expectations in the objective, but establish a feasible goal in a phased manner. Establish a purpose with foundation, you should not make a decision without prior analysis.<br>**(Key point to establish the objective priority)**<br>Define it according to the company policies. In absence of policies, analyze the importance of the problems and / or improvements. | Indicate the objective with numerical values as possible. The objective must be related to the expected effect.<br>The objective must be concrete.<br>– Example: (A specific goal should always answer a question).<br>– What? -¿ Reduce the loss of customers.<br>– How long? -¿ During 2017<br>– How much? -¿ Reduce the loss of customers by 50 |

### Work team selection

- Select key people who are directly involved and who receive benefits from the process
- Include names, roles, and responsibilities for the team within the project
- Define who the project manager is that will monitor all activities and phases from case study start to finish.

#### *4.1.2 Planning*

### Research objective

Design and implement a decision-making sales process model for a Peruvian

SME in the tourism sector, which is part of the travel management process and that allows the management and integration of unstructured data in real time. The process model is aligned with the needs currently presented by SMEs in the area under study in order to achieve the reduction of time and costs incurred in the travel management process.

**Problem domain**

A main problem for SMEs related to tourism is the excess of response times and costs associated with SME travel management processes, due to the lack of decision making in selling tourist packages according to consumer interests and market trends, in addition to generating a dependence on the companies that control it, which does not allow the generation of competitive advantages at the commercial and technological level in a context of globalization, competitiveness and greater participation in the market. Among the causes we have:

- Companies related to the tourism sector do not know the benefits of real-time data analysis technologies
- There are no models for specialized decision making in the tourism sector
- The platforms for data processing are generic and do not contemplate the particularities of information for a specific sector

**Commercial strategy to combat the problem**

Despite the fact that SMEs have limited resources, with the outlined guidelines it is possible to identify the company's current situation, with the main problem being "inadequate sales planning", which directly affects the processes of travel management in charge of defining all the travel packages throughout the year. Before identifying the company's current situation, a series of steps must be followed in order to analyze how the company has been working; identify the business processes, describe the relationship between processes and optimize and reorganize the processes, this is done prior to the development of the business architecture. As exposed in (Articulating (mis)understanding across design discipline interfaces at a design team meeting) not always the information shared in a multdisciplined group is understood. That is the reason why there must be multiple meetings with the SME's directors and user leaders in order to clarify any misunderstanding and have a clear view of the SME's current state. In order to develop this case study, it is very important to have a trustworthy contact within the SME in order to coordinate face-to-face meetings with user leaders, who has greater operational knowledge within the company [18]; in some cases these users are SME bosses or managers, but because this type of company has a brand line of command, it is usually the first option to be cited in meetings. During meetings, they must be presented to the work team or the implementing team, as well as those responsible to the company leaders, as well as to the user leaders who participate throughout the implementation, through the following steps:

- Identify business processes: to identify travel management processes, it is key to request commercial information, so that we may analyze how the

company makes its sales and how it creates travel packages, the time it takes to make a decision and what criteria it uses.
- Describe the relationship between processes: at meetings, how user leaders explain the business from their point of view should be noted, as well as recognizing the process for which they are responsible and the functions they perform. Thus we identify the current situation in the company, allowing us to recognize daily problems that affect the development of its activities. In this particular case, the time it takes to make a decision on which tourist package are created is longer than usual due to the lack of information that the company has about user interests, depending on wholesales companies to retrieve a list of pre-arranged packages and also having to wait for the user to approach the company to fill out an application, generating delays in the delivery of information.
- Optimize and reorganize the processes: with the information collected and the problem detected, a scenario is presented to improve the As-Is processes with which the company works day-to-day, thus substantially reducing the time it takes to make a decision to create a package, in addition to analyzing user interests over time, allowing packages to be generated in different seasons without having to wait for the user to approach the company.

Taking all that information into account, we can start with the business architecture to improve and propose new alternatives and reduce the problem that affects the company.

## 4.2 Phase 2: Process Design

### *4.2.1 Business Model*

The business model seeks to define the business processes according to the scope and determine how they are interrelated in order to provide value to the subsequent solution for the posed problem. It is worth mentioning that for a correct definition and process automation, the implementation team must determine a series of artifacts – mentioned below – in order to optimally contribute to the business model and automate processes that can be translated into business functions that serve as input to the next stage within the functional model.

The artifacts are defined in 4 stages:

**Analysis of the current situation**

- Functional Areas – the functional areas of the SME are those involved in the company's activities, since they are responsible for defining the objectives and goals they must achieve. Generally, a company is formed by at least 5 basic functional areas (management, marketing, sales, production, and accounting and finance), but it can be formed of many more (research, human resources, strategy, etc.) giving rise to the company's macro processes.

- Diagram of Objectives – the diagram of objectives is the graphic representation of objective distribution, from the general objective related to the SME's mission and vision, to the specific objectives, which are the objectives related to the functional areas.

**Analysis information**

- Process Map – the process map gathers the interrelation of all processes carried out by an SME, in addition it allows a global-local perspective of each process' location within the value chain; for this reason there are three process types (key, strategic and support).
- Justification Process – Objective – this is the integration of processes with the company objectives, allowing us to define the objectives of each process, for subsequent analysis in compliance with the SME's established goals.
- Definition of Stakeholders – defined as any person or entity that is concerned or affected by a company's activities; for example, the workers of that company, its shareholders, associations, mergers, civil and governmental organizations that are linked, etc.
- Description of Processes – this artifact describes the purpose of each process and its relationship with the objective defined in the previous artifacts, for their design in the next stage.
- Business Rules – the business rules describe the policies, rules, operations, definitions and restrictions present in an organization, which are of vital importance to achieve the company goals.

**Execution of the business model**

- Process Definition – the most important artifact, shaping everything related to the process, and consequently generating the definition of processes, which is an explanation of the activities that belong to the process.
- Process Matrix – Product – this artifact is vital for the following model, for it identifies all the entities, attributes and frequencies that participate through each of the diagrammed processes.
- Prioritization of Products – analyzes all the entities according to frequency, which are those that have the greatest impact and occurrence within an SME, in order to verify the delays or overestimation of times in the activities.
- Process Prioritization – the prioritization of processes as well as the prioritization of products allows us to analyze which process has the greatest recurrence within an SME, which refers to the process that consumes the most information, resources and time.
- RAM Matrix – a tool used in the process of developing the Human Resources Plan, consisting of the Organigrams and the Descriptions of Positions, defining the roles and responsibilities of all members within an SME.
- Matrix Actor – Process — this artifact identifies all actors involved in each process, and their interaction with the SME's other processes.
- Prioritization of Actors – the prioritization of actors allows us to analyze which actors have greater incurrence within the processes, both internal

and external, to verify the dependence of some and the overload or maldistribution of time in their functions.
- Process Inventory – the process inventory is the document that groups all the processes identified according to their levels (macro processes, processes, sub processes) and the type of process to which they belong (strategic, key or support).

**Business model closing**
Functional Decomposition – this artifact is the disaggregated form of the processes that are limited at the beginning of the investigation, but grouping them allows us to identify the activities that will be automated, so we can identify the business functions that will form the solution.

- Process Tab – the process tab is the document that permits us to visualize all processes, definitions, actors, entities and the functional area to which they belong.
- Portfolio of Products – after identifying the business functions, the portfolio allows us to prioritize which ones will be performed for the research solution, as well as presenting possible solutions that may directly affect a future problem.

## 4.3 Phase 3: Improvement Opportunities

### *4.3.1 Analysis of improvement opportunities*

Most opportunities for improvement require in-depth analysis, which involves working with the process' actors to identify possible problems and opportunities. Once the problem is identified, it is necessary to focus on the process to improve, define and identify the reason for the problem. To do this a diagnosis of the problem is needed, that is, the study of how the process is carried out in order to find the weak points:

- Identify the processes affected and estimate their influence.
- Identify the key process and prioritize improvement.
- Determine an indicator that allows monitoring.

Once the situation and the process have been identified, the cause of its malfunction must be analyzed. In order to analyze the causes, the Ishikawa diagram can be used to answer the following questions: What is failing? Where? And how do we know? To facilitate analysis of the possible causes it is important to classify the following groups:

- Causes that depend on the customers – these are cusomer weaknesseses, without factors that depend on the SME's workers.
- Causes that depend on the workers – these tend to be due to either lack of commitment or lack of skills.
- Causes that depend on the company – internal strategies, internal rules, disturbance of functions and responsibilities, inadequate service schedules that fail to take the needs of the population into account.

- Causes due to lack of resources and structure —the lack of structure or material resources.
- Cause due to dependence on other external companies – strategies, procedures or activities that force the company to depend on a larger one.

### *4.3.2 Data Model*

Therefore, the data model describes the data structure and their management, allowing an SME to do the following:

- Understand and communicate the data model to the company.
- Ensure the consistency and quality of the data applied in the company's processes.
- Make informed decisions through the timely availability of relevant and accurate information.

The data model is the reflection of the business model, because with the established requirements and the problems detected, the implementation team can define which components will be part of the technological solution.
As with the last model, it contains a series of artifacts that are part of the proposed solution:

- Data dictionary – recognizes all the attributes that form part of the process that must be considered for the automated tool in development.
- Logical and Physical Model – the systematic distribution where the interrelation of each attribute mentioned in the data dictionary is shown, besides showing the origin of each information.
- Analysis and selection of Cloud Computing Service – in order to identify and select an adequate Cloud Computing service for the model's implementation [19], the available tools are analyzed for evaluation and selection, according to the criteria and the company's current situation [20]. See Fig. 3 and Fig.4.

Based on the results obtained, we can see that Google Cloud Platform has a better weighted total score. It should be noted that this score is produced, to a greater extent, by the non-technical aspects being evaluated with the greatest weight and that all were met by the tool, giving it an advantage over the others.

| | | | Amazon | | Microsoft Azure | | Google Cloud Platform | | IBM Cloud | |
|---|---|---|---|---|---|---|---|---|---|---|
| Appearance | Description | Value | Importance | Balanced | Importance | Balanced | Importance | Balanced | Importance | Balanced |
| Technical | Cost of payment: payment for use | 0.15 | 10 | 1.5 | 5 | 0.75 | 10 | 1.5 | 5 | 0.75 |
| | Backups: How to deal with an information problem | 0.1 | 10 | 1 | 10 | 1 | 10 | 1 | 5 | 0.5 |
| | Types of servers: CPU's default configuration and Server RAM | 0.05 | 10 | 0.5 | 5 | 0.25 | 5 | 0.25 | 0 | 0 |
| | Types of discs: Technology in which the data is stored | 0.05 | 5 | 0.25 | 5 | 0.25 | 10 | 0.5 | 5 | 0.25 |
| | Security: Number of security certificates | 0.15 | 5 | 0.75 | 5 | 0.75 | 10 | 1.5 | 5 | 0.75 |
| | Migration of servers: the effort and the amount of time to carry out the migration | 0.1 | 10 | 1 | 5 | 0.5 | 10 | 1 | 0 | 0 |

**Fig. 3** Analysis and Cloud Computing Services selection, Appearance Technical

| | | | Amazon | | Microsoft Azure | | Google Cloud Platform | | IBM Cloud | |
|---|---|---|---|---|---|---|---|---|---|---|
| Appearance | Description | Value | Importance | Balanced | Importance | Balanced | Importance | Balanced | Importance | Balanced |
| Non-Technical | Access to low cost technical support | 0.1 | 5 | 0.5 | 5 | 0.5 | 10 | 1 | 0 | 0 |
| | Experience in the service | 0.05 | 5 | 0.5 | 5 | 0.25 | 10 | 0.5 | 5 | 0.25 |
| | Worldwide availability | 0.05 | 10 | 0.5 | 10 | 0 | 10 | 0 | 5 | 0.25 |
| | Marketplace: Number of Applications / Preconfigured systems to deploy on said platform | 0.05 | 10 | 0.5 | 10 | 0.5 | 10 | 0.5 | 0 | 0 |
| | Scalebility | 0.15 | 10 | 1.5 | 5 | 0.75 | 5 | 0.75 | 5 | 0.75 |
| | **Total** | **1** | | **8.5** | | **5.5** | | **9** | | **3.5** |

**Fig. 4** Analysis and Cloud Computing Services selection, Appearance Non-Technical

### *4.3.3 Application Model*

The application model allows us to identify each system and their relationships with the business, analyzing whether each system meets certain quality criteria with respect to the business' processes.

- Income Matrix – identifies the interaction of attributes with modules, so as to verify whether attributes are being used with the same functionality and/or evaluate the interaction of attributes with each developed module.
- Control board – defines dependency between business functions at the functional level, since this board will determine which functions should be performed in earlier phases, due to their importance in other business functions.
- Package diagram – shows the relationship between modules, associating the relation to the logical flow of processes embodied in the following modules.
- Integration diagram – artifact of the Business Architecture discipline used to identify service consumption among company software products, as well as information exchanges and dependencies between software products.

– Portfolio of entities – the portfolio of entities details all entity attributes and their definitions. These are grouped by process name. System requirements document – this document defines corporate and technical requirements for applications which make up the architecture of applications in the supply process model.

System prototypes – proposal for the solution which meets user requirements, covering their range of problems and/or necessities.

### *4.3.4 Network Model*

Technological or network architecture is made up of services proposed as part of the application model and services provided by the tool in development. Thus, we identified the services making up the network model, since it has a strong role in integration and supports the business, data and application models. Together they form the functional processes model for travel management. See Fig. 5.

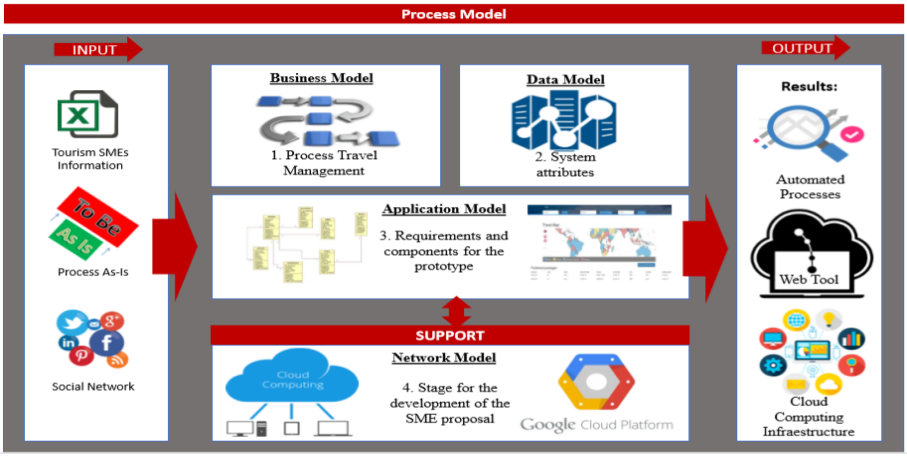


**Fig. 5** Process Model

## 4.4 Phase 4: Design and Implementation

Once the models and all the artifacts that correspond to each model are finished, a portfolio of projects is generated where the user can visualize the technological proposal, development time, solutions, benefits and all the requirements to execute it. Analyzing the proposals, four detailed components were chosen due to their low costs and production times. These are:

### *4.4.1 Cloud Platform*

Due to the current situation of tourism companies, the use of cloud services by the Google platform is projected. The services provided by Google are: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). PaaS was chosen for the case study due to the need for a platform where the solution can be developed. The other Azure, AWS, IBM Soft Layer, and Google platforms were investigated among the other cloud service platforms. Within this group Google Cloud Platform (GCP) was chosen as a viable option since it has lower operating costs than other providers and the services are user friendly. For the study case, services such as Cloud Data Store API and Machine Learning Engine were used.

### *4.4.2 Big Data – Sentiment Analysis*

The Sentiment Analysis was applied to measure tourism interests of users on the Twitter platform, which was performed by the GCP prediction API. A set of 379 examples were used as training data to be classified into four categories, good, regular, tedious and bad. This indicates with certain accuracy the feelings expressed by the user. For data extraction and analysis, we followed the steps in the following illustration. See Fig. 6.

```
Var Query = new TweetQuery();
Query.HashTags = listHashTags();
Query.Since = Yesterday();
Query.Until = Today();
Query.HasGeotag = True;
ListTweets = SearchTwitter(Query);
Foreach Tweet in ListTweets{
TweetSent = Predict(Tweet);
}
```

**Fig. 6** Pseudo-code of data loading and analysis

The following pseudo-code was used for the extraction of data from Twitter and its subsequent analysis. A query type object for Twitter is initialized, to which search parameters are assigned. In the case study, we used a list of Hashtags related to tourism, taking into account that the program will be periodically executed. The number of tweets per search was limited to those from the previous day. Finally, a flag was used to collect only the tweets containing geo-locations, to determine the origin of the tweet. The list is then sent to Google Prediction API, where a sentiment analysis was performed and the qualification returns according to prior training.

### *4.4.3 Prediction Model*

To generate a predictive model, the steps of identifying input data, analyzing available tools, implementing the model and verifying the output data are followed. This is summarized in figure 7.

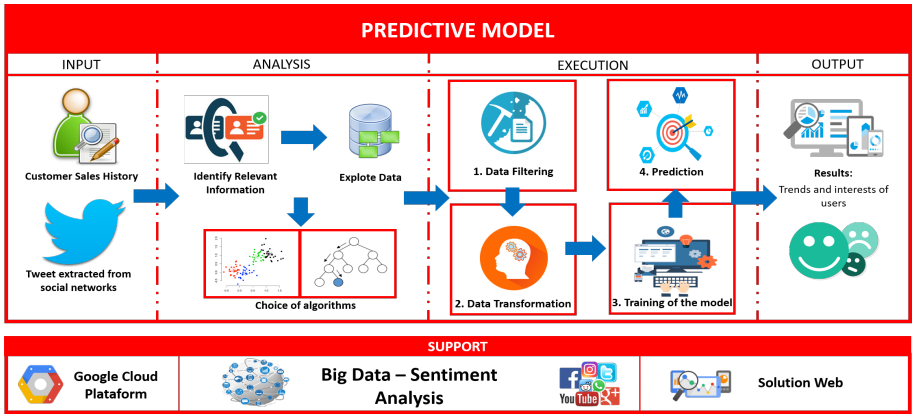


**Fig. 7** Prediction Model

A Deep Feed-Forward neuronal network was used to obtain predictions of tourist interest. The network takes the input data in the first layer, after classifying them by the hidden layers of the network, and returns a defined result in the model planning. This model uses sentiment analysis to obtain a more accurate prediction. Wide and deep models were joined, giving the neural network a greater capacity for abstraction. For the horizontal layers, characteristics such as tweet publication day or country of origin are taken, as well as cross characteristics, giving the model more certainty. The input columns were chosen for vertical features. These models are linked to DNN * (Deep Neural Network) linear regression data output. This gives us greater certainty and calculation speed.

## 4.5 Phase 5: Project Closure

It is necessary that business user acceptance tests be carried out as part ofthe control and closure phase, to conclude with model implementation. Tests must be validated by a Key User of each process taking part in the Travel Management process and post production. Approximately a one month support period must be given in order to ensure implementation success.
Once the implementation process is completed, it is recommended that SMEs have an IT staff to perform support functions such as: Access to developed

tools, backup management, improvements in functionality and incident attention.

## 5 Validation

To validate the process model presented, we applied a case study to show that the proposal successfully solves the needs of companies in the tourism sector. It must be said that it is validated as a part of the model in order to ensure its correct functioning in the case study. The components involved were the four models and their integration was used for practical validation through the case study. In addition, for confidentiality purposes, the real company will be represented as OT S.A.C.

### 5.1 OT S.A.C. Company Information

OT S.A.C is a small business that sells and distributes tour packages, with ten employees and a monthly turnover of approximately $29,000. It is located in the district of Santiago de Surco, in the city of Lima, Peru. Its main suppliers are wholesale companies within the sector, from whom it receives a list of packages for sale and distribution. In turn, this company sells tailor-made packages as required by the client.
From its early stages, according to national regulations regarding tourist package distributors, this company was favored by the portfolio of clients it already had because of related dealings. As new companies appeared, competing in the same category and with the same products, a price competition began, spurring a sudden growth in this sector. Under these circumstances, the owners and managers focused on the vision of using new technologies and trends to keep company sales afloat. That is why we proposed a process model aimed at emerging technologies to this company, in collaboration with the workers of the same, as well as involving the direct managers of the company.

### 5.2 Proposed Model Implementation

The scheme to implement the proposed model can be seen in figure 8, which details the steps that the company OT S.A.C must implement.

#### *5.2.1 Face-face Meeting*

During implementation, we held weekly meetings with the owners and workers involved, to explain what their functions consisted of, allowing them to design a process map for OT SAC, see Fig 9. We found that the main problem was within the travel management process.

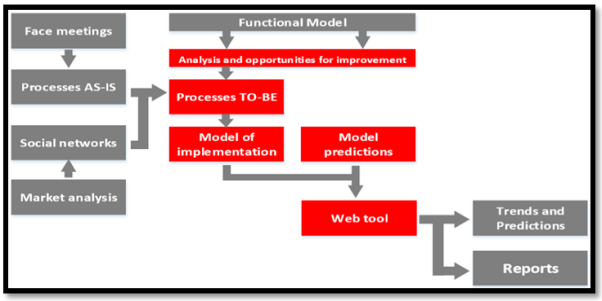


**Fig. 8** Steps to implement the model

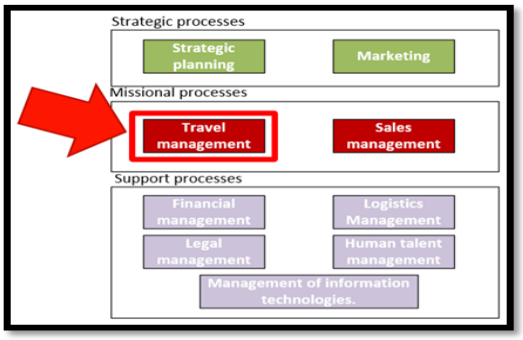


**Fig. 9** OT S.A.C Business Processes Map

### *5.2.2 Market Analysis*

We carried out two types of survey (companies and customers) to analyze the current situation regarding SMEs, in order to know customer behavior, the current situation of companies and their needs.

### *5.2.3 Analysis of SME companies*

We performed market research to study the behavior of 50 SME companies in the tourism sector, asking a series of questions that demonstrated the companies' current situation, recurrent problems and needs for future improvemnt.

- What are your main distribution channels to inform customers of package promotions? Email with 43%, Website with 27%, Newspapers 19% and social networks only 11%.
- Do you keep a log of customer purchases? Only 60% of companies keep track of customer purchases.
- Is it considered important to create your own packages without depending on wholesale companies? Only 53% of companies are happy with the current operation of their business.

- Are new technologies to analyze users trends on social media known to the company? Only 5% of companies have a notion of new technologies and the impact they can have on their companies.
- Would a tool to help analyze market trends and support new tour package planning be accepted in the company? 47% were willing to try new tools that support the planning of their tour packages.

As a result, companies use e-mail as the main distribution channel and only 5% understand the benefits of being able to study customers on social networks.

#### *5.2.4 Customer analysis*

We conducted a client investigation to answer the following question: "What is the most used information search channel, and why?"; we conducted a survey with a sample of 300 users distributed across four well-defined age groups (17-24, 25-40, 41-60, and 60+). The group with the highest percentage (40%) was 25 to 40 years old. Gender distribution was 55% male and 45% female. In the surveys carried out by the research team, interviewees said they used the following search channels: Social Networks, Travel and Tourism Companies, Friends Recommendations, Newspapers and magazines, other.
The main aim of this survey was to inquire about travel trends and user interest; we found that 46% use social networks as a search channel before planning any trip, and found that the main reason was the degree of reliability in users who share favorable or negative experiences during their trips. Therefore, companies in the tourism sector may use social networks as a new channel to collect information and make decisions based on users needs and trends.

#### *5.2.5 As-Is processes*

Regarding the information survey, OT SAC announced that the company did not have well-defined processes, so we were unable to design an As-Is process; however we noted all the functions performed by each worker to analyze their activities and later the processes that they carried out.

- In the As-Is processes elaboration, company dependence on the wholesalers that control the market becomes clear, and in the market analysis a new information collection channel is presented, where companies can make decisions based on users' needs and trends shown on social networks.

#### *5.2.6 As-Is processes*

To analyze and propose improvement opportunities, we constructed a cause-effect diagram to identify the causes of the main problem and the improvements that can be applied. See fig. 10 The cause and effect analysis presents the main problems as travel package definition strategies, not knowing client interests, poorly defined promotion dissemination strategies, and poor package reservation planning, generating difficulties for the majority of companies.

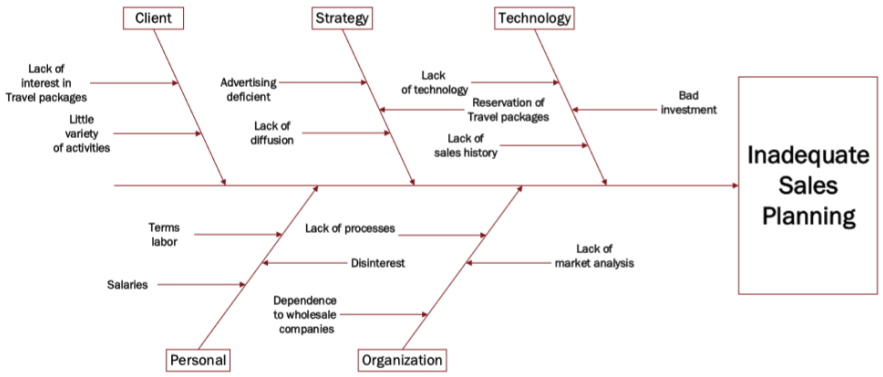


**Fig. 10** Diagram causes - effect

Thus we propose a design for new information search channels aiming to reduce poor sales planning and understand client interests in depth, to define travel packages according to customer needs.

### *5.2.7 To-Be processes*

With the company processes and causes of the problem identified, a meticulous work was carried out to define the processes that the company should implement for optimal performance, and without gaps that could affect its profitability, taking the functional model as foundation, in order to reference the components and thus be able to compare their information gaps.

- In this particular case, the following is identified within the Travel Management process:
- The process does not have its own documentation or any means of information other than from companies that control the market.
- Their information input is not in real time but from a history provided by other companies.
- The response time to solve user requirements is limited to the response provided by another company that has the information.

These are the main problems with the functional model. To solve these findings, we implemented a new information search channel within the process, supported by an application developed for the company, thus allowing the company to make decisions in real time according to the user trends and predictions the tool gives us. This new process was created so that the company can work with the new application implemented. See Fig. 11.

### *5.2.8 Functional decomposition*

In the new search channel process, the most important business functions for the program are detected, such as the analysis of customer sales histories, the

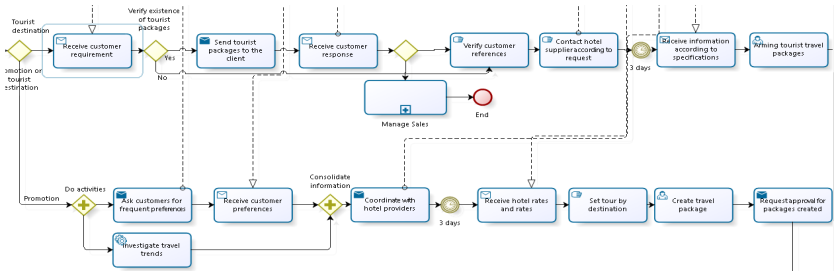


**Fig. 11** New information search channel

analysis of sales requests and social network data mining. These address the operation of the web application, which is why we proposed the OT S.A.C process model.

### *5.2.9 Prototype operation*

For the development of a web tool, data mining from social media is proposed as a research objective. For the case study, Twitter was chosen. This tool will analyze user comments and identify their preferences, tastes, and mood, inter alia. This will be reflected in the qualification of the city, country and tourism package. In addition, this information will be taken as a data source to later predict which tourist destinations have greater acceptance. Sentiment analysis concepts were used to analyze the relevant information and make a better decision in the creation of tour packages.

### *5.2.10 Prototype implementation*

Process model validation was carried out in an SME company in the tourism sector, for this purpose and as part of the research, all four modules were integrated in the following phases.

**Start phase:**
The first step for model implementation is to evaluate whether the company meets the minimum requirements for model adaptation, which are: that only the processes of tourist destination management and destination promotion management are part of the Travel Management process.

**Planning Phase:**
This phase consists in planning the migration of the SME's current travel management process model to the proposed model, for which a project team must be established, which must consist of the following roles: Project manager, Functional analyst, Programmer analyst and Test Analyst.

**Execution Phase:**
In the execution phase, the model must be adopted in the developed web tool; the following tasks will be carried out by each member of the implementation team:

- Functional Analyst: Responsible for loading the process diagrams, data model and organizational structure.
- Programmer Analyst: Responsible for loading the developed services, in addition to adding the necessary functionalities to meet the requirements and solve the problems detected.
- Test Analyst: In charge of performing tests on the processes executed and their adequate interaction with requested services, as well as the information load in the database.

**Control and Closing Phase:**
Once the implementation process is completed, it is recommended that the SME have an IT staff in order to perform support functions such as: access the web tool, backup management, improvements in functionality and incident attention.

### *5.2.11 Predictions Model*

After analyzing the benefits of using a predictive model in tandem with sentiment analysis, the solution was implemented in OT S.A.C. For optimal performance, the predictive model was trained with the output data from the sentiment analysis. When the neural network is used to infer a result, at each step it learns and increases to the certainty of the result. The tool shows evidence of learning for each iteration.

## 5.3 Web tool

For prediction model construction, we developed a web tool to collect Twitter comments, in order to analyze them and identify the user's preferences. Using Sentiment Analysis concepts, were identified fields such as activity type, geolocation, publication dates and finally the essential part; the tweet. See Fig. 12.

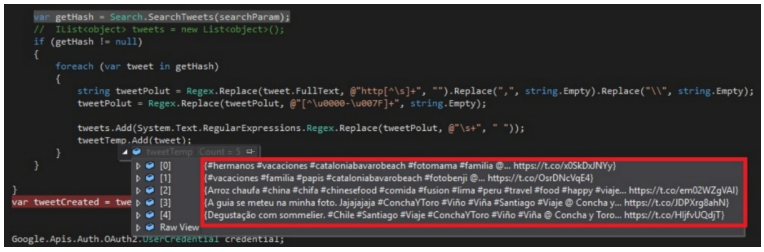


**Fig. 12** Snip of the tweet pool

Using these results and the analysis of the integrated predictions model, the tool can accruately predict user trends and activities, see Fig.13. It can also

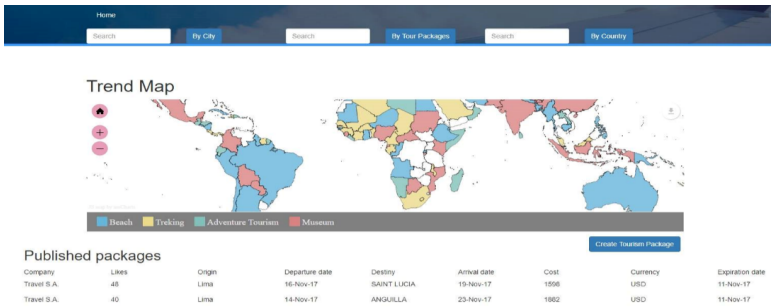


**Fig. 13** User tendency map

identify the days that users rate better, since there is a metric that analyzes this information. Added to this is the visualization of packages that users have found suitable for them.

## 5.4 Results

At the end of the process model implementation in OT SAC, we can see that the benefits and indicators in the processes using new trends and technologies ensure a greater number of sales, considering that with mass data analysis tools, according to International Data Corporation [21], there was a progressive sales increase of 12%. Hereto, we analyzed the delays in sales packages delivery based on user queries; according to the sample of 300 users, the average time was about 13 days, ranging from 10-15 days.
After process model implementation and thanks to the support of the web tool, OT SAC reduced decision-making time by 60%, with an average of 6 days (range 3-8 days).

On the other hand, thanks to the support provided by the forecasting model, OT SAC was able to reduce logistic expenditures. In addition to increasing new travel packages sales, they were able to create their own packages, reducing dependence on a wholesale company. See Fig.14.
Regarding the results, we can say the following: Before implementing the process model, the company lost customers for not responding to time requirements, and lost package reservations due to not considering customers' needs. However, after implementation, not only can the need be recognized, but also the company is able to negotiate new rates with suppliers that will generate a greater return on sales.
According to the feasibility analysis (see Fig. 15), investment and recovery are increased after implementation. It shows a 12-month investment of $24,212.92. That includes both manpower for implementation and cloud services, after that period the only cost incurred would be for the cloud services, with an estimated profit of $25,874.36 from secured sales. These latter values are made possible by an investment return of approximately 12 months, with an ROI of

| Sales comparison | | | |
|---|---|---|---|
| **Before** | | **After** | |
| Separation of Package with Wholesaler. | 10% | Separation of Package with Wholesaler. | 0% |
| Coordination with transport. | 5% | Coordination with transport. | 5% |
| Average sales | 6 | Average sales | 13 |
| Total price | $ 2810 | Total price | $ 2810 |
| Gain per package(20%-10%) | $ 281 | Gain per package(20%) | $ 562 |
| **Total gain** | **$1686** | **Total gain** | **$ 7384** |

**Fig. 14** Before and after sales comparison.

114%, which is very profitable for a company in the tourism sector. This leads us to conclude that the process model, with the web tool support, would be a benefit for the sector as it generates a new information channel which could create new opportunities in the market.

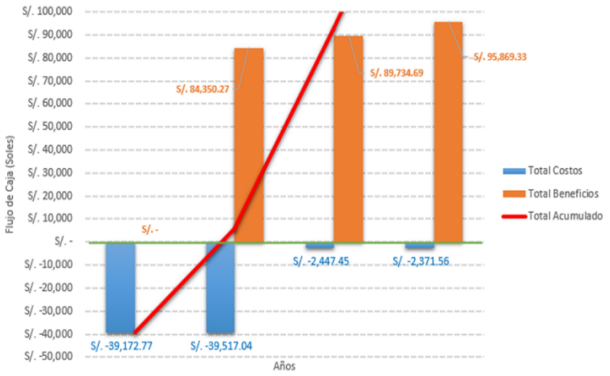


**Fig. 15** Feasibility study.

## 6 Conclusions and Future Work

In the present study, we demonstrated that a possible solution to the problem of package sales decision making is a process model in tandem with a support application by means of process re-engineering and Cloud Computing. All this is of the outmost importance for SMEs in order to obtain both financial and operational stability.

We demonstrated that the use of technologies such as Cloud Computing oriented to Free Software guideline can be profitable for the SME sector, given

that the cost of these services is within the purchasing power of small businesses.

It was possible to demonstrate that an IT vision increases profitability to an SME indirectly, saving time and money. It also provides the means to take immediate action in the face of an unexpected situation. For future research, it is possible not only to continue with the extension of the model, but also to exploit the tools and expand the analysis to all social networks in order to have more accurate and possibly new information markets.

## 7 Recommendations

It is recommended for the analysis phase to carry out a survey with the SMEs in person, to identify the SMEs' problems, needs and requirements, considering that the survey must be carried out with the staff that execute the processes in the SMEs. These meetings and agreements must be documented in minutes.

It is recommended to have a confidentiality of information document in order to use the SME's information in the implemented model, as established in an agreement signed by both parties. It is recommended that when implementing the model, online support – either personalized or from software documentation – may be required, considering its integration with the other components of the model. It is recommended to have a project continuity proposal. Likewise, it is necessary for the SME to formally accept the implementation, in order to leave evidence that the implementation was successful.

## 8 Acknowledgments

This work has been partially funded by the following projects of the Spanish Ministry of Science, Innovation and Universities GROMA (MTM2015-63710-P), MODAS-IN (reference: RTI2018-094269-B-I00), PPI (RTC-2015-3580-7) and UNIKO (RTC-2015-3521-7), and the "methaodos.org" research group at URJC.

## References

1. Polastri, G.(May 4, 2015). Google Perú: Solo el 15% de pymes peruanas utilizan herramientas digitales en su negocio. Gestión. Retrieved from https://gestion.pe/economia/google-peru-15-pymes-peruanas-utilizan-herramientas-digitales-negocio-88525
2. Garimella, K., De Francisci Morales, G., Gionis, A. and Mathioudakis, M. (2016). Quantifying Controversy in Social Media. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (pp. 33–42). New York, NY, USA: ACM. https://doi.org/10.1145/2835776.2835792

3. Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J.M., and Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments, Annals of Operations Research, 236(1), 197–213. https://doi.org/10.1007/s10479-013-1449-6
4. Baucom, E., Sanjari, A., Liu, X., and Chen, M. (2013). Mirroring the real world in social media. In Proceedings of the 2013 international workshop on Mining unstructured big data using natural language processing - UnstructureNLP '13 (pp. 61–68). New York, New York, USA: ACM Press. https://doi.org/10.1145/2513549.2513559
5. Eckerson, W. W. (2006). Predictive Analytics. Extending the Value of Your Data Warehousing Investment.TDWI BEST PRACTICES REPORT. Retrieved from http://www.sas.com/events/cm/174390/assets/102892_0107.pdf
6. Dubin, R. (2008). Theory building. Free Press. Retrieved from https://books.google.com.pe/books?id=a0NqAAAAMAAJ
7. Chauhan, R., and Kaur, H. (2013). Predictive Analytics and Data Mining. Business Intelligence, 359–374. https://doi.org/10.4018/978-1-4666-9562-7.ch019
8. Zapata, G., Murga, J., Raymundo, C., Alvarez, J., and Dominguez, F. (2017). Predictive Model based on Sentiment Analysis for Peruvian SMEs in the Sustainable Tourist Sector. In Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (pp. 232–240). SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0006583302320240
9. Huang, L., Luo, Q., and Zou, C. (2013). A Framework of New Rural E-Government and the Related Information Resources Integration. International Journal of u- and e- Service, Science and Technology, 6(6), 83-96. https://doi.org/10.14257/ijunesst.2013.6.6.09
10. Lu, Y., Deng, Z., and Wang, B. (2007). Tourism and Travel Electronic Commerce in China. Electron. Market., 17(2), 101–112. https://doi.org/10.1080/10196780701295974
11. Yan, W. (2014). Research on Tourism E-Commerce Based on Web Data Mining Technology. Applied Mechanics and Materials, 543–547(Iea), 3663–3666. https://doi.org/10.4028/www.scientific.net/AMM.543-547.3663
12. Shao, H., Zhang, Y. and Li, W. (2017). Extraction and analysis of city's tourism districts based on social media data. Computers, Environment and Urban Systems, 65, 66–78. https://doi.org/10.1016/j.compenvurbsys.2017.04.010
13. Brandt, T., Bendler, J. and Neumann, D. (2017). Social media analytics and value creation in urban smart tourism ecosystems. Information and Management, 54(6), 703–713. https://doi.org/10.1016/j.im.2017.01.004
14. Zeng, Z. Y.and Ye, X. Y. (2013). Research on Jiangxi Tourism E-Commerce Platform Construction. Advanced Materials Research, 734–737, 2990–2993. https://doi.org/10.4028/www.scientific.net/AMR.734-737.2990
15. Zhang, X. M. and Xu, C. (2012). A Semantic E-Tourism Framework Based on Cloud-Computing. Applied Mechanics and Materials, 197, 661–668. https://doi.org/10.4028/www.scientific.net/AMM.197.661
16. Höpken, W., Fuchs, M., Keil, D. and Lexhagen, M. (2015). Business intelligence for cross-process knowledge extraction at tourism destinations. Information Technology and Tourism, 15(2), 101–130. https://doi.org/10.1007/s40558-015-0023-2
17. Merida, C., Rios, R., Kobayashi, A. and Raymundo, C. (2017). Modelo arquitectónico de información para una plataforma de Big Data para el Sector Turístico. In Proceedings of the Décima Sexta Conferencia Iberoamericana en Sistemas, Cibernética e Informática (CISCI 2017), Decimo Cuarto Simposium Iberoamericano en Educacion, Cibernetica e Informatica, SIECI 2017 - Memorias, pp. 321-328.
18. Luck, R. (2013). Articulating (mis)understanding across design discipline interfaces at a design team meeting. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 27(02), pp.155-166. http://doi.org/10.1017/S089006041300005X
19. Sikeridis, D., Papapanagiotou, I., Rimal, B. P. and Devetsikiotis, M. (2017). A Comparative Taxonomy and Survey of Public Cloud Infrastructure Vendors, 1–20. Retrieved from http://arxiv.org/abs/1710.01476
20. Harris, T. (2010). Cloud Computing Services – A comparison. Retrieved from http://www.thbs.com/downloads/Comparison-of-Cloud-computing-services.pdf
21. IDC. (2016). Worldwide Semiannual Big Data and Analytics Spending Guide. Retrieved from https://www.idc.com/getdoc.jsp?containerId=prUS41826116