Algerian republic democratic and popular

Ministry of Higher Education and Scientific Research

Larbi Tébessi university –Tébessa-

Faculty of Exact and Natural Sciences and Life

Department: Mathematics and computer science

Master thesis

Domain : Computer science

Division: Computer science

Option: Information Systems

Title:

Intelligent Decision Support System based on Big data for urban planning

Presented by:

KHEDIRI Abderrazak

In front the jury:

H.BENDJENNA	Dr	**Larbi Tébessi University**	**President**
M.R.LAOUAR	Pr	**Larbi Tébessi University**	**Supervisor**
A.ZEBDI	A.P	**Larbi Tébessi University**	**Examiner**

Graduation date: 29-05-2016

Mark : …. **Mention : ………………………..**

Dedication

For you my father, God rest your soul, you're gone but you're still alive in my heart and you'll stay forever, rest in peace dear father...

For you my dear mother, the person who providing me the climate, and bring me every day the balance without which I could never keep advancing, God bless you.

For my sisters and brother.

For my family

For my friends.

# Acknowledgment

I want to thank my supervisor, Professor LAOUAR M.R who guided and directed this work.

I want to thank also Dr BRADJI L for his encouragement.

My thanks go to the LAMIS Laboratory staff.

My thanks also to my teachers.

I would like thank everyone who helped me in my studies.

Abstract

Decision making in spatial and geographical domain in general, and urban planning in particular become progressively more difficult and may have detrimental consequences. Due to the multi-criteria, uncertainty and complex nature of spatial issues, traditional Decision Support Systems are no longer able to manage the process and support Decision Makers. Recent years, many methods, models and systems for classification has been developed, each of them has its own advantages and drawbacks, they are insufficient to meet the complexity of these problems, so new tools and techniques are required.

In this thesis, the presented work aims mainly to elaborate an architecture also a prototype of intelligent Decision Support System that provides relevant assistance to urban planners in urban projects. Our work allows facilities during the process of urban planning, knowing the public desires and tendencies by integrate citizens in the process to know their degree of satisfaction, and take into account the storage of the large amount of urban data (Big Data) that increase with exponential velocity.

The proposed work treat the field choice problem integrate many techniques that contribute the intelligence in urban decision making. Basing on previous models (fields) and using naïve Bayes classifier which is a Data mining technique, we make prediction of the selected strategy consequences, classify the alternatives according to the decision maker desire, the outcomes are used to increase the performance. Using clustering we categorize citizens to ensure the quality of participants and try satisfy them. Also we thought about a Big data tool that can support the quantity of data which is HBase. Finally a prototype has been developed to concretize our proposition.

Key words: Urban planning, intelligent Decision support system, Data mining techniques, Prediction, Naïve Bayes, Clustering, Big data.

Résumé

La prise de décision dans le domaine spatial en général, et de la planification urbaine en particulier devient progressivement plus difficile et peut avoir des conséquences néfastes. En raison des multicritères, l'incertitude et la complexité des questions spatiales, les systèmes d'aide à la décision traditionnelle ne sont plus en mesure de gérer le processus et le soutien des décideurs. Ces dernières années, de nombreuses méthodes, modèles et systèmes de classification a été mis au point, chacun d'eux a ses propres avantages et inconvénients, ils sont insuffisants pour répondre à la complexité de ces problèmes, de sorte que de nouveaux outils et techniques sont nécessaires.

Dans ce mémoire, le travail présenté vise principalement à élaborer une architecture aussi un prototype un système d'aide à la décision qui fournit une assistance pertinente pour les planificateurs urbains dans les projets urbains. Notre travail offre des facilités énormes au cours du processus de la planification urbaine, de connaitre les désirs et les tendances par l'intégration des citoyens dans le processus ainsi que connaître leur degré de satisfaction et de prendre en compte le stockage de la grande quantité de données urbaines (Big Data) cette augmentation avec une vitesse exponentielle.

Le travail proposé sert à traiter le problème de choix de terrain en intégrant de nombreuses techniques qui contribuent l'intelligence dans le processus décisionnel urbain. Basant sur les modèles précédents (champs) et en utilisant naïve classificateur Bayes qui est une technique de fouille de données, nous faisons la prédiction de la décision pour connaitre les conséquences de la stratégie choisie, classer les alternatives selon le désir décideur, les résultats sont utilisés pour augmenter la performance. En utilisant le clustering nous catégorisons les citoyens pour assurer la qualité des participants ainsi que les satisfaire. Aussi, nous avons pensé à un outil de Big data qui peut supporter la quantité de données qui est HBase. Enfin, un prototype a été développé pour concrétiser notre proposition.

Mots clés: La planification urbaine, système intelligent d'aide à la décision, Techniques de data mining, prédiction, Naïve Bayes, Clustering, Big data.

ملخص

ان اتخاذ القرارات في المجال المكاني والجغرافي بشكل عام، والتخطيط الحضري على وجه الخصوص أصبح تدريجيا أكثر صعوبة، وأنه قد يكون له عواقب ضارة. ويرجع ذلك إلى معايير متعددة منها الطبيعة المعقدة للمسائل المكانية ونظم دعم اتخاذ القرار التقليدية التي لم تعد قادرة على إدارة هذه العملية ودعم صناع القرار. في السنوات الأخيرة وضعت العديد من الأساليب، التقنيات، النماذج ونظم التصنيف وكل منهم لديه مزايا وعيوب الخاصة به، إلا أنها غير كافية لتلبية تعقيد هذه المشاكل، لذلك هناك حاجة إلى أدوات وتقنيات جديدة.

في هذه الأطروحة، يهدف العمل المقدم أساسا لوضع استراتيجية ونموذجا أوليا لنظام ذكي لدعم اتخاذ القرار الذي يقدم المساعدة ذات الصلة لمخططي المدن في المشاريع الحضرية. عملنا يقدم عديد التسهيلات خلال عملية التخطيط الحضري معرفة رغبات العامة وميولهم، بعد دمجهم في عملية التخطيط لمعرفة درجة رضاهم وكذلك اخذ بعين الاعتبار تخزين كمية كبيرة من البيانات الحضرية التي تزداد بسرعة هائلة.

العمل المقترح يصبو الى علاج مشكلة اختيار حقل وذلك من خلال دمج العديد من التقنيات التي تساهم بذكاء في عملية صنع القرار في المناطق الحضرية. مستندة على النماذج السابقة (الحقول)، وباستخدام المصنف بايز البدائي والذي يعتبر تقنية للتنقيب عن البيانات، نتوقع عواقب الاستراتيجية المختارة وتصنيف البدائل وفقا لرغبة صانع القرار وتستخدم النتائج لرفع مستوى الأداء. وأيضا باستخدام تقنية التجميع نصنف المواطنين لضمان نوعية المشاركين ومحاولة إرضائهم. كما فكرنا في أداة للبيانات الكبيرة التي يمكن أن تدعم كمية البيانات وأخيرا تم تطوير نموذج أولي لتجسيد افتراضنا.

Table list

Figure List

Contents

Chapter 1 Context and problematic: Urban planning

Chapter 2 Decision Making: State of the art

Chapter 3 Proposed architecture of IDSS for urban planning

Chapter 4 Implementation and experimentation of the proposed architecture

General introduction

Formerly, people lived and worked mainly in rural areas, nowadays the world has changed, it has undergone reforms and transformations in all domains and especially in the urban domain, the speed with which was carried out and in particular in developing countries, coupled with the inefficient management of urbanization by the states concerned, led to sometimes chaotic urban situations.

Populations around the world are largely urban. More than half of them live in urban areas, including population growth has accelerated, which has led consequences: new aerated towns and cities growth like mushrooms, public authorities should apply solutions that can be realized: Either a new development, an enlargement of the city or even rehabilitation. All these solutions within the scope of urban planning.

Urban planning is one of the most complex domains that you could make decision in. because of including many actors and the multi-criteria nature of spatial problems. Several conflicting criteria are often taken into account in evaluating different Planning processes, which present a problem of agreeing a solution that satisfy all participants. In the other hand there is projects that seems good for urban planners but, the citizens express their dissatisfaction.

In addition to the huge quantity of data which increase rapidly, and which can't be managed by frozen standard tools. They are insufficient to meet the complexity of these problems, so new tools and techniques are required.

Therefore traditional system are no longer able to assist policymakers. Consequently we propose in this thesis architecture of intelligent system that have the ability to solve the cited problems and offers better performance.

Structure:

Chapter 1:

In this chapter we define what is urban? What is an urban project? Sustainable development, urban management, urban planning, the problems related to the urban context and the issue to be treated in this thesis.

Chapter 2

We talk about several topics in this chapter which was mainly the decision support especially spatial domain, DSSs and IDSS Also related works for each domain of them.

Chapter 3

We discuss in this chapter our contribution which was composed of an architecture and there principal components and processes, the demarche of such process.

Chapter 4

We present in this chapter our proposed prototype and the used data relative to the future system we have applied a use case consisting of real collected data.

Chapter 1

Context and Problematic: Urban planning

1 Introduction

Urban planning is the set of tools and resources aims make the city healthier, bigger and nicer to provide a safe, organized, and enjoyable home and work life for citizens. We can said that is a skill that involves social choices and requires methods. It is based on a prospective and strategic vision that takes into account the potential and physical constraints, social, economic and environmental territory. In this chapter we spreads the fundamental concepts related to urban context.

2 Urban concept

2.1 What we mean by urban?

Urban can be defined as follows *"is a place-based characteristic that incorporates elements of population density, social and economic organization and the transformation of the natural environment into a built environment"* [WEEKS, 2010]

2.2 Urban or urbane?

Most people know an urban place from the first sight. Although defining it is not as easy as it might seem. [Weeks, 08] defined urban as being a characteristic of place, rather than of people. Places are typically defined as "urban" and on the basis of that definition the people living there are thought of as being part of the urban population but, we do not usually apply the term "urban" to a person. The personal adjective "**urbane**," still occasionally used to describe a person, is defined by the Oxford English Dictionary as ***"having the qualities or characteristics associated with town or city life; esp. elegant and refined in manners, courteous, suave, and sophisticated"***

2.3 What is an urban project?

Traditional notions plan and planning are gradually replaced by those of sustainable urban development and urban project. The urban project has several dimensions, and can be defined as follows: *"The urban project is both a collaborative process and territorial project: it consists to define and implement management measures on an urban territory given, partnership with all civil and institutional partners concerned, integrating the different territorial levels and long term, for sustainable urban development "*.[VILLEDURABLE, 2014]

3 Urban and territory plan-making

Designing and implementing plans is something very rigorous, concerning its content and also the tasks to be accomplished and project management. [LAURINI, 02]

3.1 Urban planning definition

Because of sensibility of this section we don't modify a lot of expression. We bred for the specialists and professionals to talk about it.

According to [DOUBRIERE, 79] [LAURINI, 02] urban planning must make the city healthier, bigger and nicer, together with safeguarding and showing to advantage the city's heritage. But [HENDERSON, 97] is convinced that no single definition can possibly cover the breadth of information and practice that makes up contemporary planning. For him, urban planning is best thought of as a process that uses a variety of tools (zoning, transportation planning, environmental policy, housing programs, etc.) to achieve envisioned and desired goals within the natural and built environments. There are varieties of professional activities that may be considered planning, including *land use planning*, *transportation planning*, *environmental planning*, *social planning* and *urban design*. Each of these has different underlying theories, technical practices, and prominent leaders. However, most planning activities have the following four qualities in common [**LAURINI, 02**]:

1. Planning is future-oriented. The decisions made in the planning process are generally made to affect a future condition in the environment. This seems obvious, but it is a good thing to reinforce.

2. Planning is concerned with defining and evaluating alternative solutions to problems. This is deeply rooted in rational planning theories that underlie current planning practice. It is based on the premise that it is difficult to argue that a chosen strategy is the right one to pursue if alternatives have not been defined for comparison and evaluation.

3. Planning is political. Every public planning decision takes place in a political context. It is important to realize early on that the majority of planning activities involve the use or regulation of land in some form, and that all land belongs to someone who is afforded rights, most importantly the right to due compensation.

4. **Planning has a special** responsibility to represent the needs of minorities, the disabled, the poor and other underrepresented groups. Planners are required to pay special attention to the needs of these groups as part of their professional code of ethics. [**LAURINI, 02**]

3.2 Territory development

In case of using the adjective *territorial* before the word development it almost refers to either the *spatial integration* or the *geographic scale*, of development.

According to [**Romeo, 2015**] territorial development have two senses: the first is technical and the second is neutral

In the technical sense, territorial development *"refers to integrated multi-sector development across a specific portion of territory, guided by a spatial vision of the desirable future* *and supported by strategic investments in physical infrastructure and environmental management. This definition makes no reference to scales (local, regional, national or transnational) and applies equally to any of them."*

In the neutral, sense, territorial development is simply *"an umbrella term for development of specific (typically sub-national) portions of territory. These may be an urban, metropolitan, regional or rural jurisdiction, but also watershed, coastal, mountainous, border areas, etc. Most often,"* he said also that the term is used to encompass both *local* development (narrowly associated with smaller, first-tier, jurisdictions or even part of them and regional development, i.e. the development of larger, intermediate jurisdictions (districts, provinces, regions, etc.). [Romeo, 2015] continue with saying that the expression *territorial* development may just designate local development at *any* scale, since any space can be defined as local from an observer placed above or outside it.

[Romeo, 2015] answer the question of what designate? Territorial development designates development that is endogenous and spatially integrated, leverages the contribution of actors operating at multiple scales and brings incremental value to national development efforts.

3.3 Sustainable development

[Butlin, 89] defined it as *"development that meets the needs and aspirations of the present without compromising the ability of future generations to meet their own needs"*. Thus, is the organizing principle for sustaining finite resources necessary to provide for the needs of future generations of life on the planet. It is a process that envisions a desirable future state for human societies in which living conditions and resource-use continue to meet human needs without undermining the "integrity, stability and beauty" of natural biotic systems.

3.4 Urban management

Urban management means: [**LAURINI, 02**]

- Data collected on a daily basis to foster municipal information systems (accounting, building permits, etc.)
- Different kinds of administrative authorizations
- Maintenance of the city infrastructures
- Social services

For an elected official, urban planning reflects a mid-term or a long-term vision of the city, whereas the urban monitoring or management system is for the short term.

Urban management, there are many connections between urban planning and urban management. Among others, urban planning assumes and implies that daily management is correctly done; and if not an objective of urban planning can be to correct this.

3.5 Urban planning versus urban management of sustainable cities

The fundamental purpose of urban planning is to create new plans taking into consideration the necessary objectives that can be taken as a gradual change in phase of the plan. Urban management is made to follow the gradual change and plans. In this section, we will try to make some comparisons between urban planning and urban management. After that, some elements for planning and management of sustainable cities will be given in order to emphasize the connections between short-term and long-term consequences.

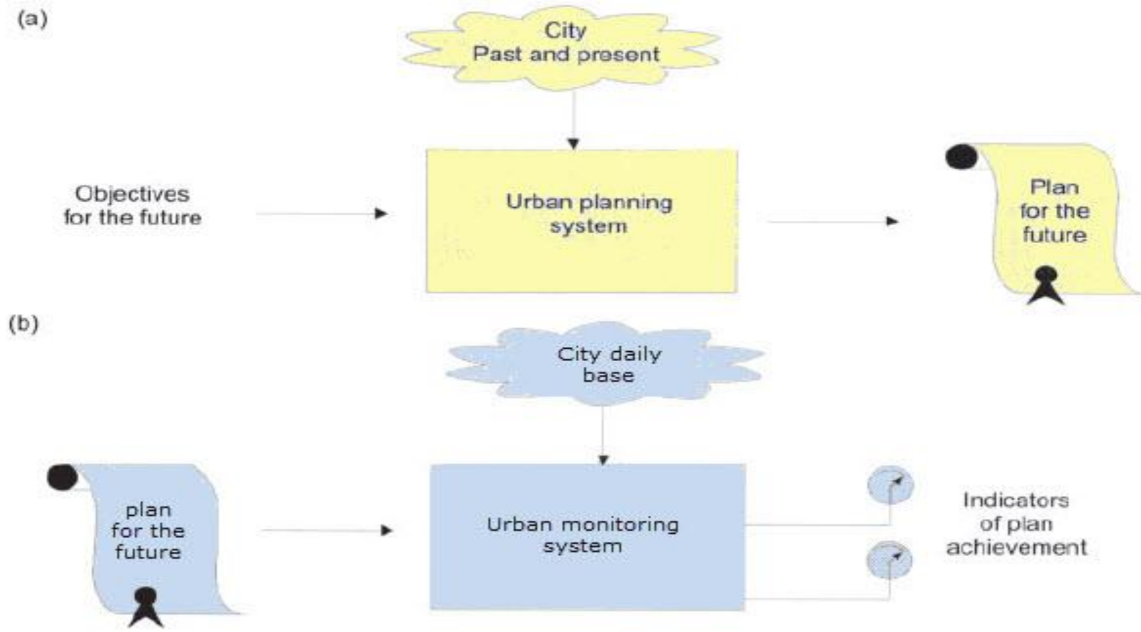


Fig 1.1 Comparing systems for (a) urban planning, (b) urban monitoring. [**LAURINI, 02**]

4 Urban problem

4.1 Urban problem definition

According to [HAYES, 81] a problem can be defined as follows: *"Whenever there is a gap between where you are now and where you want to be, and you don't know how to find a way across that gap, you have a problem"*. He continues by defining solving as *"finding an appropriate way to cross that gap"*.

4.2 Problems related to the urban context

Several questions arise here, as the definition of urban problem, their particularities and other questions. Declaring that urban problem have reasons and does not exist by itself as theoretical answer, because it is only the facet of a more general societal problem [CASTELLS, 77]. No more information about its nature is giving from this debate-expanding.

[Chadwick, 71] was saying:

Problem=goal+impediment to the goal

This definition was extended by [Faludi, 73] declaring that *'a problem is a state of tension between the ends pursued by a subject and his image of the environment'*. So, three aspects can be pointed out:

- The importance of the actor (person, group, organization)
- The perception of reality
- The possible future.

Because urban problems identification represent the huge difficulty, and the [Castells, 77] answers is just theoretical, let's see a little more deeply and clearly what urban problems are. According to several urban actors it is not worthwhile to develop sophisticated instruments to identify them; some major actors prefer waiting to see the problem set by some other actors.

For instance **[LAURINI, 02]** examine perfectly the noise pollution problem in cities and we will take it as it is:

If you look at claims sent by residents to the city hall, it is very interesting to see that:

- No complaints are sent from people living in the vicinity of very noisy motorways
- The majority of complaints come from very quiet residential areas.

If a city planner were to develop a strategy to deal with urban noise by using only the number of complaints received, the results would be very strange; he would plan to reduce noise in very quiet areas and do nothing in very noisy areas. However, more or less, all urban problems are linked together to form a sort of mess. (Fig 1.2)

To identify solutions represent the main issue. Urban problem solving is a very difficult challenge, perhaps the most difficult our society has presently to face. Alas, very often, by trying to solve a single problem, some new ones can emerge.

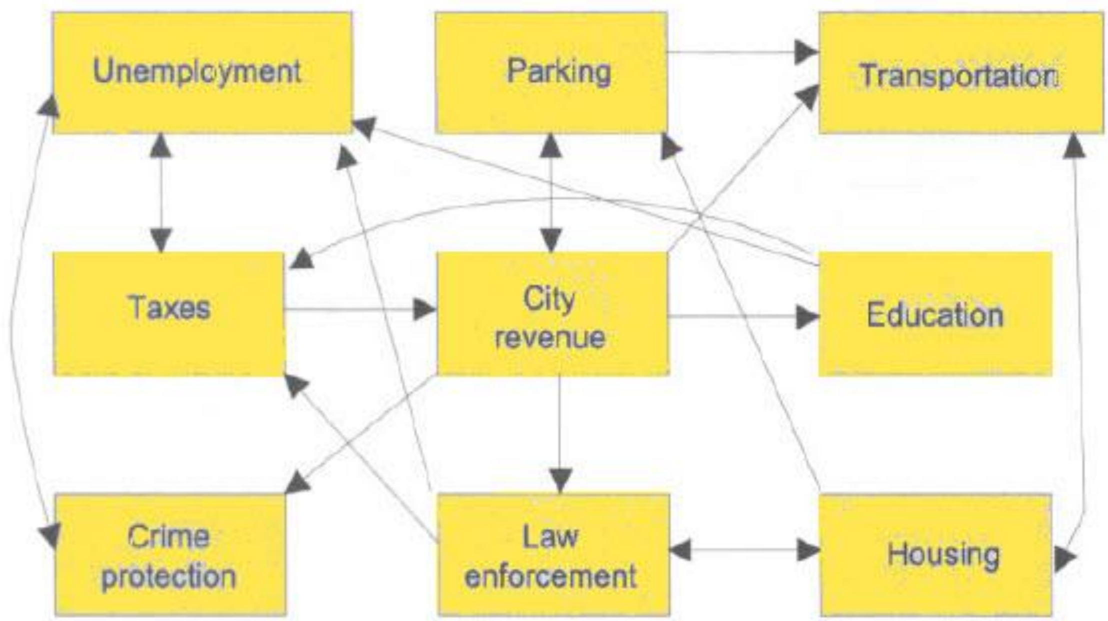


FIG 1.2 the mess of urban problems [**LAURINI, 02**]

5 Planning for sustainable cities

5.1.1 Fundamental principles of sustainable development

The concept of sustainability represent one of the main concerns of urban planning. Sustainability is defined by the World Commission on Environment and Development

[WCED, 87] as the *"development that meets the needs of present generations without compromising the ability of future generations to meet their needs and aspirations"*. According to [MAY & AL, 96], the fundamental principles of sustainable development are as follows:

• **Futurity** (also known as inter-generational equity or 'not stealing from our children and grandchildren')

• **Social equity** (also known as intra-generational equity or 'care for today's poor and disadvantaged')

• **Environment** (environmental conservation and protection).

• Public participation, in other words, individuals should have an opportunity to participate in decisions that affect them and in the process of sustainable development

5.2 Model of a sustainable city

Having a decent quality of life for both present and future citizens is the aim of sustainability.in the other hand quality of life must represent a balance or a compromise between a number of issues, considering physical and mental health of persons as well as the situation of the community and public participation. All those issues can be integrated into a model the general framework of which can be seen in Fig 1.4.

In order to reach those objectives, one of the possibilities is to develop a model, or more exactly a set of integrated sub-models, of a sustainable city.

According to [MAY ET AL, 96], any model should always:

- avoid nebulous relationships
- develop sub-models for particular places, where equations can be specified with more confidence

- emphasize problems which are difficult to manage, have complex causes and wide-ranging impacts

- enable a contrasting range of policies to be tested.

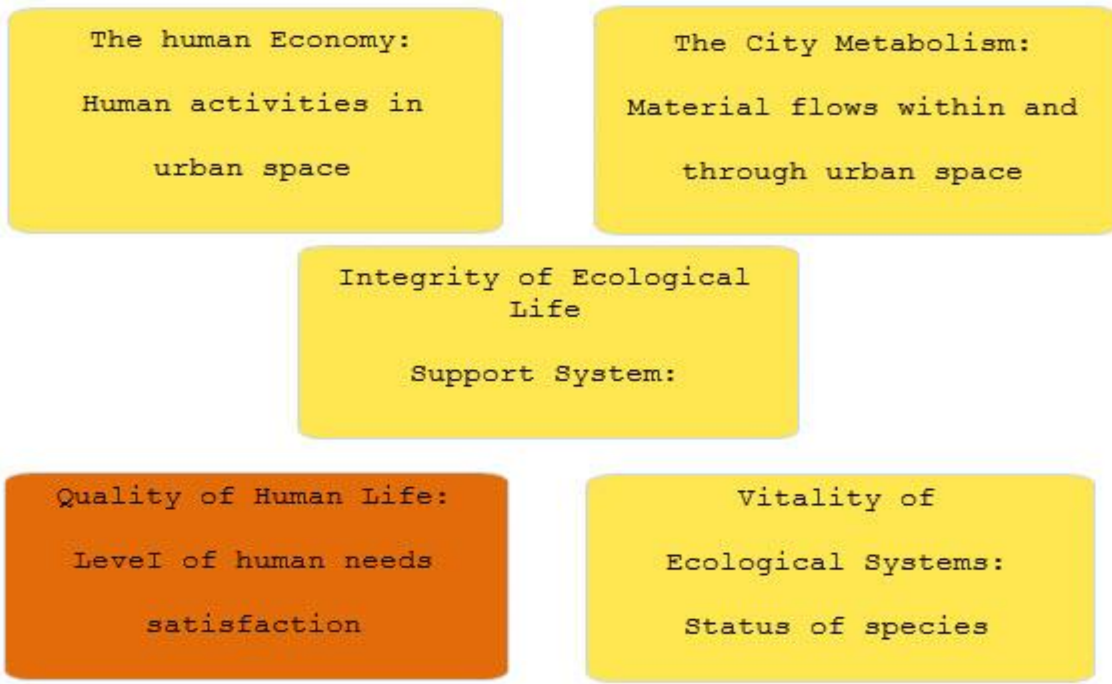


FIG 1.3 Components of a conceptual framework for a sustainable city model.

[May et AL, 96]

6 Problematic

The Algerian population is now largely urban. More than half of Algerians live in urban areas, including population growth has accelerated, which has led many any negative consequences than positive: new aerated towns and cities are mushrooming in this case, public authorities should apply solutions. The Solutions that can be achieve are: Either a new development, an enlargement of the city or even a rehabilitation.

All these solutions fit into the context of urban planning. At each planning exercise they need a field to realize the project. The participating party (Administrations) must give their opinions on proposed land. Where each of them has its easement (see annex for servitude). They vary according to the administration (Public Works Direction, Agriculture Direction, Town Planning Department, Building and Housing ...). Each field has its own characteristics such as the nature of land (agricultural, clay...), driving under the ground-(of electricity cables, telephone cables, gas lines ...). Planning the group must decide which field will host the project from the proposals.

This decision regardless of good or bad, remains in force for twenty or thirty years. The obvious delay of urban development today; is due to bad decision taken, and the dissatisfaction of the citizens of realized projects which represent one the fundamental principles of sustainable development which also remains a big problem.

So the group must make a decision that meet both them and the citizens, the problem is that in Algeria the public authorities suffer from a share of a fatal shortage in planning tools. Furthermore the lack of intersectoral cooperation sometimes arises the problem of obsolescence of tools and other hand they remain relatively static.

7 Conclusion

In this chapter we tried to define as precisely as possible. What is urban? What is an urban project? Sustainable development, urban management, urban planning, the problems related to the urban context and the issue to be treated in this thesis.

The next chapter describes the decision support in general, the Decision support systems, the intelligence as well as the means used to make intelligent systems in such a state of the art.

Chapter 2

Decision Support: State of the art

1. Introduction

When it comes to solving a simple problem, the decision is made by reflection experiment based on the decision maker. Although it is a process or a decision-making process to follow, so the presence of a model is needed.

The majority of human activities requires taking daily decisions; whether at the level of a country, region, an administration, a local community, a business within the family or just the scale of the individual.

To better understand the scope of the application of decision support methods to urban issues and the need to improve the conventional methods, it worth a brief review of the evolution of this field of study.

Policy makers are turning more and more towards the Geographic Information Systems (GIS) to help solve complex spatial problems. However, these systems do not adequately take into account the decision support because they lack analytical modeling capabilities. The answer to these shortcomings is the development of Intelligent Spatial Decision Support Systems (ISDSS), specifically designed to meet in spatial problems. The design of these systems has seen a leap forward in recent decades to become very powerful tools. This chapter aims to give an overview on the basics of Decision support in general and multicriteria decision support, intelligent decision support and the different used techniques in particular.

2 Decision support (Decision aiding)

According to [ROY, 96] Decision aiding can be defined as:" *the activity of the person who, through the use of explicit but not necessarily completely formalized models, helps* *obtain elements of responses to the questions posed by a stakeholder in a decision process. These elements work towards clarifying the decision and usually towards recommending. Or simply favoring a behavior that will increase the consistency between the evolution of the process and this stakeholder's objectives and value system".*

Decision Support is dedicated to simplify and assist in constructing, establishing, and arguing for convictions. The basis and the means of developing the decision must be the object of critical discussion [ROY, 96]

2.1 Decision

We can consider the decision as the action when a single individual (the decision maker) exercises freely choose between several options of shares at some point. [LESOURNE ET AL, 02] affirm that «deciding is what you do when you don't know what to do ». This is the traditional picture of a decision maker making a decision that is closing or, etymologically, «cutting off» the matter. When we do not know what to do, we either make a decision or we put it off until later [LESOURNE & AL, 02]

2.2 Decision process

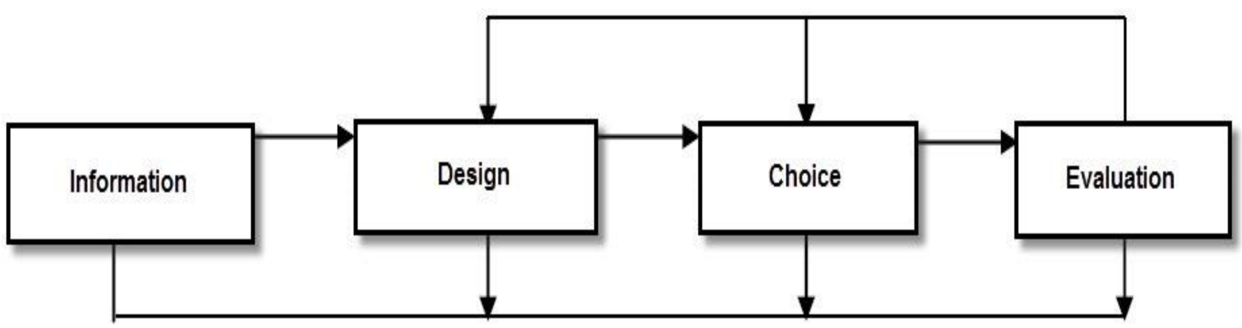


Fig 2.1: Simon's Decision process [Simon, 77]

The Information: This is the phase that determines all the necessary data (but not necessarily sufficient) that will be used in subsequent phases

The Design: This phase generates the different alternatives that make all the possibilities. Various solutions are developed at this stage

The choice: This is the phase of restricting all possibilities subset of selected options

Assessment (Evaluation): In light of the previous three phases of the solution Temporarily retained as satisfactory, this phase can lead to Reactivation of one of the three previous stages or, conversely, to validate the solution.

2.3 Decision typologies

Every decision has a certain inherent degree of structure about it [Roy, 96].

[Keen & Morton, 78] differentiate **structured** (programmed) **and unstructured** decisions in the following manner:

In unstructured decisions, the human decision maker must provide judgment and evaluation as well as insights into the problem definition. They also distinguish perceived structure from deep structure and contend that it is important for IS researchers to consider perceived structure especially in the context of DSS design and development. This is important mainly because in organizational decision making, the context plays a significant role in determining both the decision process and the contents

[Roy, 96] use the notion of programmed decision and unprogrammed (non-programmed) decision, other authors called them the decision styles.

Structured decisions	**Unstructured decisions**
Routine, repetitive	Unexpected, infrequent
Established & stable contexts	Emergent & turbulent contexts
Alternatives clear	Alternatives unclear
Implications of alternatives straightforward	Implications of alternatives indeterminate
Criteria for choosing well defined	Criteria for choosing ambiguous
Specific knowledge needs known	Specific knowledge needs unknown
Needed knowledge readily available	Needed knowledge unavailable
Result from specialized strategies (i.e., procedures that explicitly pre-specify full set of steps to follow in order to reach decisions)	Result from general strategies (e.g., analogy, lateral thinking, brainstorming, synthesis used in the course of reaching decisions)
Reliance on tradition	Reliance on exploration, creativity, insight, Ingenuity

Tab 2.1 Decision structuredness [Burstein & Holsapple, 08]

2.4 Decision maker

The decision maker is a main contributor (entity) which plays a central role in the decision process and **to whom** the decision support.

The decision maker, then, is the one who assesses the "possible" and the objectives, and who expresses preferences and has an interest in imposing them on the evolution of the process [ROY, 96]

The actors (individuals, entities, communities) are what we shall call **stakeholders** [ROY, 96]

Different actors can have different styles of decision making. Some are naturally inclined to make decisions in a systematic manner. On the other hand, there are decision makers who are comfortable with using mainly their intuition. The importance of decision making style of the decision maker can hardly be exaggerated. [KULKARNI & AL, 07]

2.5 The Analyst

We will going here to answer the question: **aiding by whom?**

The decision maker may not have the background to perform the decision aid. The one performing the aid is, therefore, generally different from the decision maker. Whether distinct from the decision maker or not, we shall call this individual **the analyst**. [ROY, 96]

2.6 Decision of group

When the decision include many individuals and entities (decision maker) not only a single individual, we are here in front to decision of group. Before the decision making the group can realizes two types of activities: negociation and participation. So there is a negociation based decision and participation based decision.

2.6.1 Negociation:

The activity or the process of discussing something with someone or group of individual in order to reach an agreement with them, or the discussions themselves. For example we can say the agreement of the site that will host the University was reached after a series of difficult negotiations. The exact details of the agreement are still under negotiation.

For decision making involving groups with more divergent objectives, the final decision is likely to involve some form of political negotiation between stakeholders, each of whom may adopt different sets of criteria for evaluating alternatives. [BELTON & STEWART, 02]

2.6.2 Participation:

Cambridge dictionary defines the participation concept as "*the fact that you take part or become involved in something*".

The participatory approach consist to develop a culture of participation. This concept is founded on a voluntary basis

Different definitions have been proposed for the concept of participation. We will retain that of [METTAN, 92]

"All *democratic opportunities for the population to collaborate in decision making. Authentic participation is then the active involvement of citizens in solving the problems they consider essential and relevant they can compete effectively control solutions"*

2.7 Multiple Criteria Decision Aiding (MCDA)

As shown by its acronym MCDA (Multiple Criteria Decision Aiding/Analysis) is a DSS which the adopted tool of analysis is a multi-criteria method.

One of the principal aims of MCDA approaches is to help decision makers organise and synthesize such information in a way which leads them to feel comfortable and confident about making a decision, minimizing the potential for post-decision regret by being satisfied that all criteria or factors have properly been taken into account. [BELTON & STEWART, 02]

Thus, "we use the expression MCDA as an umbrella term to describe a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter." [BELTON & STEWART, 02]

2.8 Typologies of problem, for which MCDA may be useful

[ROY, 96] identifies four different problematiques, i.e. broad typologies or categories of problem, for which MCDA may be useful:

The choice problematique: To make a simple choice from a set of alternatives.

The sorting problematique: To sort actions into classes or categories, such as "definitely acceptable", "possibly acceptable but needing more information", and "definitely unacceptable".

The ranking problematique: To place actions in some form of preference ordering which might not necessarily be complete

The description problematic: To describe actions and their consequences in a formalized and systematic manner, so that decision makers can evaluate these actions. Our understanding of this problematique is that it is essentially a **learning problematique**, in which the decision maker seeks simply to gain greater understanding of what may or may not be achievable.

In [BELTON & STEWART, 02] the authers added other problematiques:

The design problematique: To search for, identify or create new decision alternatives to meet the goals and aspirations revealed through the MCDA process, much as described by [KEENEY ,1992] as "value focused thinking".[BELTON & STEWART, 02]

The portfolio problematique: To choose a subset of alternatives from a larger set of possibilities, taking account not only of the characteristics of the individual alternatives, but also of the manner in which they interact and of positive and negative synergies. [BELTON & STEWART, 02]

2.9 Multiple criteria decision support methods

The multiple-criteria decision support is a major study branch of operational research involving several schools of thought, mainly American with the work of Thomas L. Saaty who invent Analytical Hierarchy Process method (AHP) and European with those of Bernard Roy in LAMSADE[1] Laboratory who invent ELECTREI, ELECTREII and ELECTREIII methods.

These methods and calculations to choose the best or the optimal solution among a range of solutions or eliminate some solutions from the total solutions.

The MULTIPLE CRITERIA DECISION SUPPORT METHODS ALSO CALLING THEM outranking approaches differ from the value function approaches in that there is no underlying aggregative value function. The output of an analysis is not a value for each alternative, but an outranking relation on the set of alternatives.

An alternative a is said to outrank another alternative b if, taking account of all available information regarding the problem and the decision maker's preferences, there is a

LAMSADE [1]: Laboratoire d'analyse et modélisation de systèmes pour l'aide à la décision strong enough argument to support a conclusion that *a* is at least as good as *b* and no strong argument to the contrary. The way in which the outranking relation is exploited by a method depends on the particular problematique (choice, sorting, or ranking…). [Belton & Stewart, 02]

Or simply we can say: "an action in outperforms (outrank) another " if it is at least as good as the other, with respect to a majority criteria, without being too significantly worse than the other relative to the rest of the criteria.

2.9.1 ELECTRE

ELECTREI, ELECTREII, ELECTREIII, ELECTRE IV and ELECTRE Tri are family of multi-criteria decision analysis methods that originated in Europe in the mid-1960s. The acronym ELECTRE[2] stands for: ELimination and Choice Expressing REality.

The family of ELECTRE methods use the outranking approach (*relation de surclassement* in French), it differ according to the degree of complexity, or richness of the information required or according to the nature of the underlying problem or problematique.

The ELECTRE methods are based on the evaluation of two indices, namely the *concordance* index and the *discordance* index, defined for each pair of options *a* and *b*. The concordance index, *C(a, b),* measures the strength of support in the information given, for the hypothesis that *a* is at least as good *b.* The discordance index, *D(a, b),* measures the strength of evidence against this hypothesis. [Belton & Stewart, 02]

ELECTRE[2] : Elimination Et Choix Traduisant la REalité (Fr)

2.9.2 PROMETHEE

PROMETHEE (Preference Ranking Organization MeTHod for Enrichment Evaluations) is a family of methods to help the developed multi-criteria decision. The PROMETHEE I (partial ranking) and PROMETHEE II (complete ranking) were developed by J.P. Brans and presented for the first time in 1982, they have undergone many changes at the initiative of the authors (Jean-Pierre Bertrand Brans and Mareschal) of the Free University of Brussels. **[BRANS & MARESCHAL, 05]**

PROMETHEE is a prescriptive approach to multi-criteria analysis problem with a number of decisions evaluated according to several criteria. It is associated with the descriptive approach, to visualize the conflicts and synergies between criteria, GAIA (geometrical analysis for interactive aid).

The main feature of the PROMETHEE methods is that each possible extension will be very clear and easily understood by the decision-maker. **[BRANS & VINCKEJ, 85]**

Many interactive softwares such as Decision Lab 2000 D-Sight and PROMETHEE can use the PROMETHEE methods.

2.9.3 Analytical Hierarchy Process (AHP)

AHP is a method of multi-criteria decision support and a structured technique based on mathematics and psychology for organizing and analyzing complex decisions considering several criteria in order to make the best decision. It has been created in the 1970s by Thomas Saaty, professor at the Wharton School of Business and a consultant to the US government and has been extensively studied and refined since then. He developed AHP to optimize the allocation of resources when there are several criteria to consider a complex decision may be based on dozens of criteria for decisions and potential solutions.

AHP facilitates analysis solutions by structuring them hierarchically using your criteria.

It has particular application in group decision making [Saaty, 08], and is used around the world in a wide variety of decision situations, in fields such as government, business, industry, healthcare, shipbuilding [Saracoglu, 2013] and education.

In addition of the capability of prescribing "correct" decision, the AHP helps decision makers find the one that best fits their understanding of the problem and their own goal. It provides and offers a comprehensive and rational scope for many purpose: structuring a decision problem, representing and quantifying its elements, relating those elements to overall goals, and evaluating alternative solutions. [JAKŠIĆ & RAKOČEVIĆ, 2012]

The Analytic Hierarchy Process (AHP) approach, does not escaping the rule, it also uses the outranking relationship for example, makes direct use of such intuitive statements, by allowing decision makers to give verbal descriptions of relative importance in terms such as "moderately", "strongly" or "absolutely" more important, which are converted into assumed ratios. It seems possible that one of the reasons for the popularity of AHP (and the associated Expert Choice software) is in fact the natural appeal of such semantic scales for purposes of expressing relative importance. [Belton & Stewart, 02]

3 Decision Support Systems

We have already defined the decision support, multiple-criteria decision analysis (aiding) and their methods moving now to systems. The concept of a decision support system (DSS) is extremely broad and its definitions vary depending on the author's point of view (Druzdzel and Flynn 1999). What is a Decision Support System?

A Decision Support System *"Is a computer-based information system that supports business or organizational decision-making activities. DSSs serve the management, operations, and planning levels of an organization and help to make decisions. Decision support systems can be either fully computerized, human or a combination of both"*. [Power, 07]

Other author defined DSS as "a system under the control of one or more decision makers that assists in the activity of decision making by providing an organized set of tools intended to impart structure to portions of the decision-making situation and to improve the ultimate effectiveness of the decision outcome". [Marakas, 99]

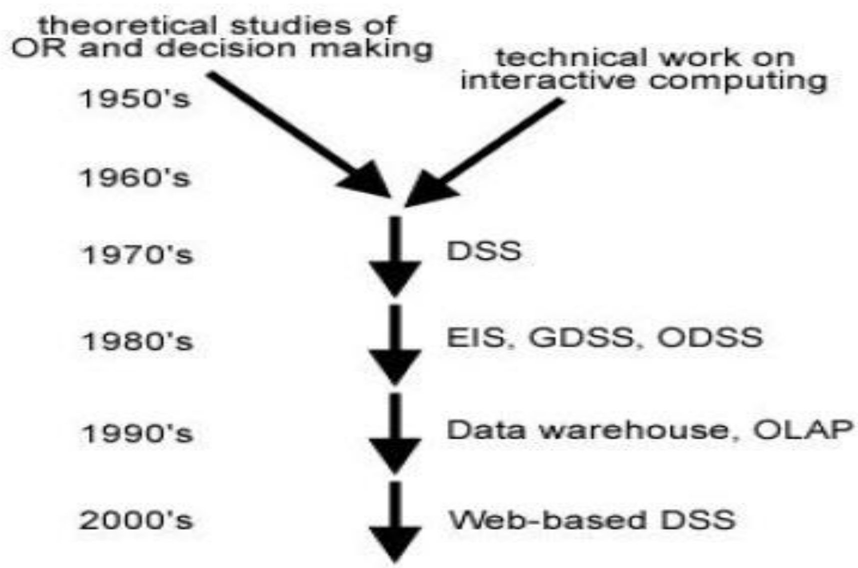


Fig 2.2: A brief history of DSS

3.1 DSS taxonomies

Same case of definition, there is no taxonomy that included any DSS. Different authors propose different classifications. Hättenschwiler (1999) [Hättenschwiler, 01] differentiates *passive*, *active* and *cooperative DSS*.

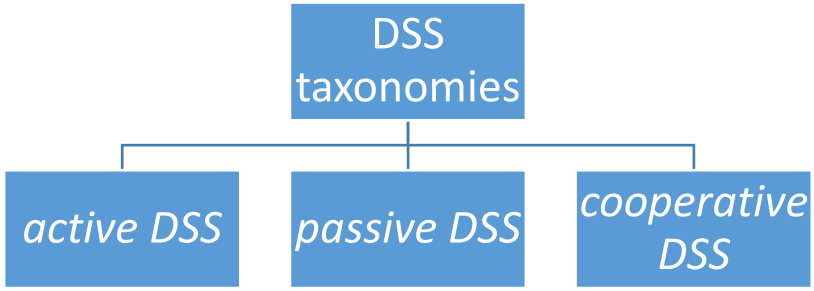


Fig 2.3 DSS taxonomies according to [HÄTTENSCHWILER, 01]

An ***active DSS*** can bring out such decision suggestions or solutions.

A ***passive DSS*** is a system that aids the process of decision making, but that cannot bring out explicit decision suggestions or solutions.

A ***cooperative DSS*** allows the decision maker (or its advisor) to modify, complete, or refine the decision suggestions provided by the system, before sending them back to the system for validation. The system again improves, completes, and refines the suggestions of the decision maker and sends them back to her for validation. The whole process then starts again, until a consolidated solution is generated. [HAETTENSCHWILER, 01]

At the conceptual level, Power (2002) differentiates *Communication-Driven DSS, Data-Driven*

DSS, Document-Driven DSS, Knowledge-Driven DSS, and Model-Driven DSS.

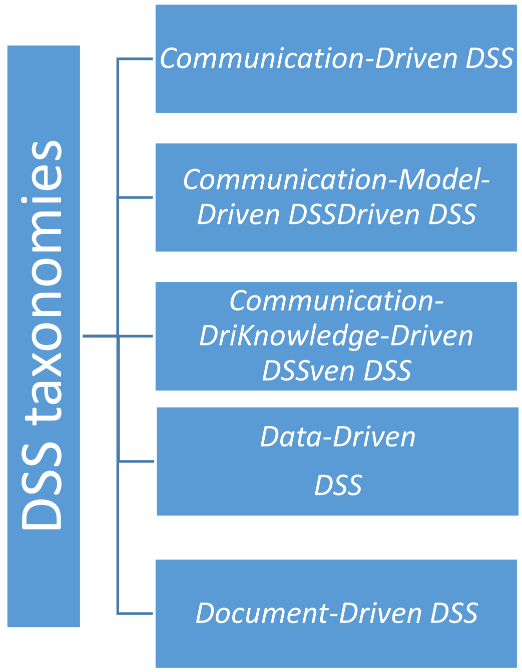


Fig 2.4 DSS taxonomies according to **[Power, 02]**

A ***Model-Driven DSS*** emphasizes access to and manipulation of a statistical, financial, optimization, or simulation model. Model-Driven DSS use data and parameters provided by DSS users to aid decision makers in analyzing a situation, but they are not necessarily data intensive.

A ***Communication-Driven DSS*** supports more than one person working on a shared task; examples include integrated tools like Microsoft's NetMeeting or Groove (Stanhope 2002). ***Data-Driven DSS*** or ***Data-oriented DSS*** emphasize access to and manipulation of a time-series of internal company data and, sometimes, external data.

Document-Driven DSS manage, retrieve and manipulate unstructured information in a variety of electronic formats.

Finally, ***Knowledge-Driven DSS*** provide specialized problem-solving expertise stored as facts, rules, procedures, or in similar structures.

3.2 DSS architecture

Many authors identify different components in DSS [Sprague & Carlson, 82], [Hättenschwiler, 01], [Power, 02]. Based on the various existing architectures, [Marakas, 99] proposes a generalized architecture made of five distinct parts: *(a)* the data management system, *(b)* the model management system, *(c)* the knowledge engine, *(d)* the user interface, and *(e)* the user(s)

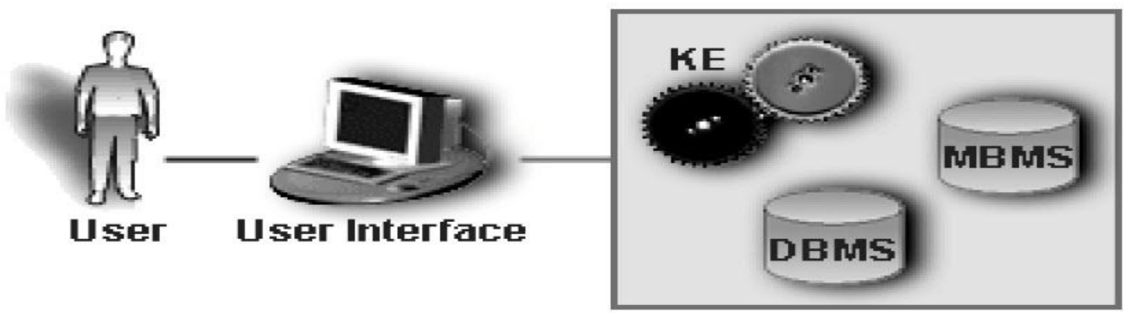


Fig 2.5 DSS architecture [Marakas, 99]

3.3 Spatial decision support system

A spatial decision support system (SDSS) *"is an interactive, computer-based system designed to support a user or a group of users in achieving a higher effectiveness of decision making while solving a semi-structured spatial decision problem"* [Malczewski, 99]. There are in the intersection of two major predisposition in the field of spatial science: science of geographic information (GIS) and spatial analysis. the enormous level of complexity of spatial problems and the particular nature of data (geographic) considered in different spatial problems (urban, territorial, rural…) are the reasons that make SDSS different of traditional decision support system (DSS).[Chakhar & Mousseau, 07].

3.4 Relationships decision-makers, the territory DSS

After talking about the spatial decision support systems, decision makers and the territory (chapter 1). What are the relations between them?

[LAURINI, 02] present a schema which shows the different relationships between them.

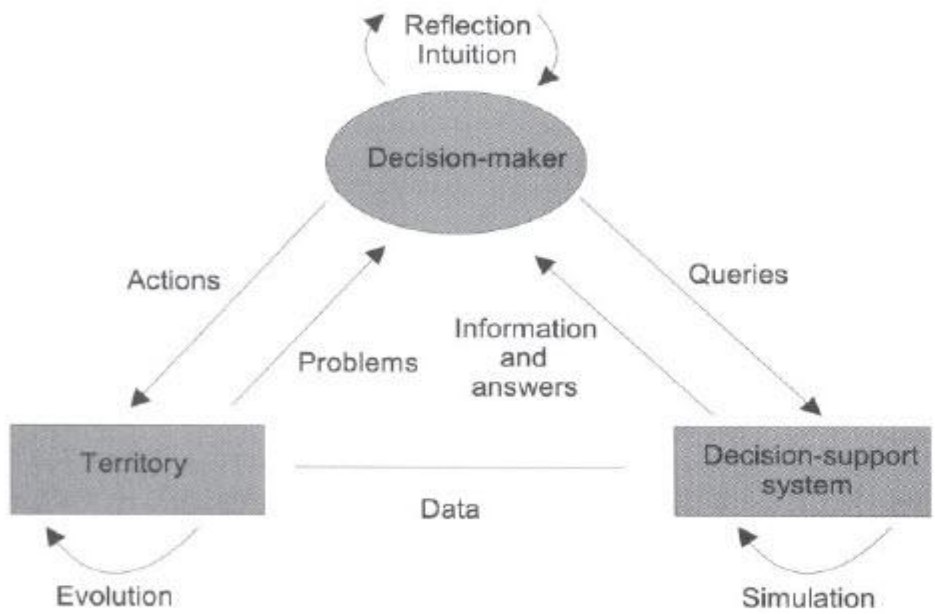


Fig 2.6: Relationships between decision-makers, the territory and the DSS **[LAURINI, 02]**

3.5 Multicriteria spatial decision support systems (MC-SDSS)

Multicriteria spatial decision support systems (MC-SDSS) can be considered as part of the wider field of SDSS. The specificity of MC-SDSS is that it aids, assists and supports multicriteria decision making in spatial domain. Spatial multicriteria decision making refers to the use of multicriteria analysis (MCA) to solve spatial decision problems. **[FIGUEIRA ET AL, 05]** **[CHAKHAR & MOUSSEAU, 07]**

3.5.1 Related work

[Chakhar & Martel, 03] Proposed a Framework also a design (Architecture) (Fig 2.7) conceived of in such a way that it supports GIS-MCA integration and is also open to incorporate any other OR/MS tool into the GIS as the author say.

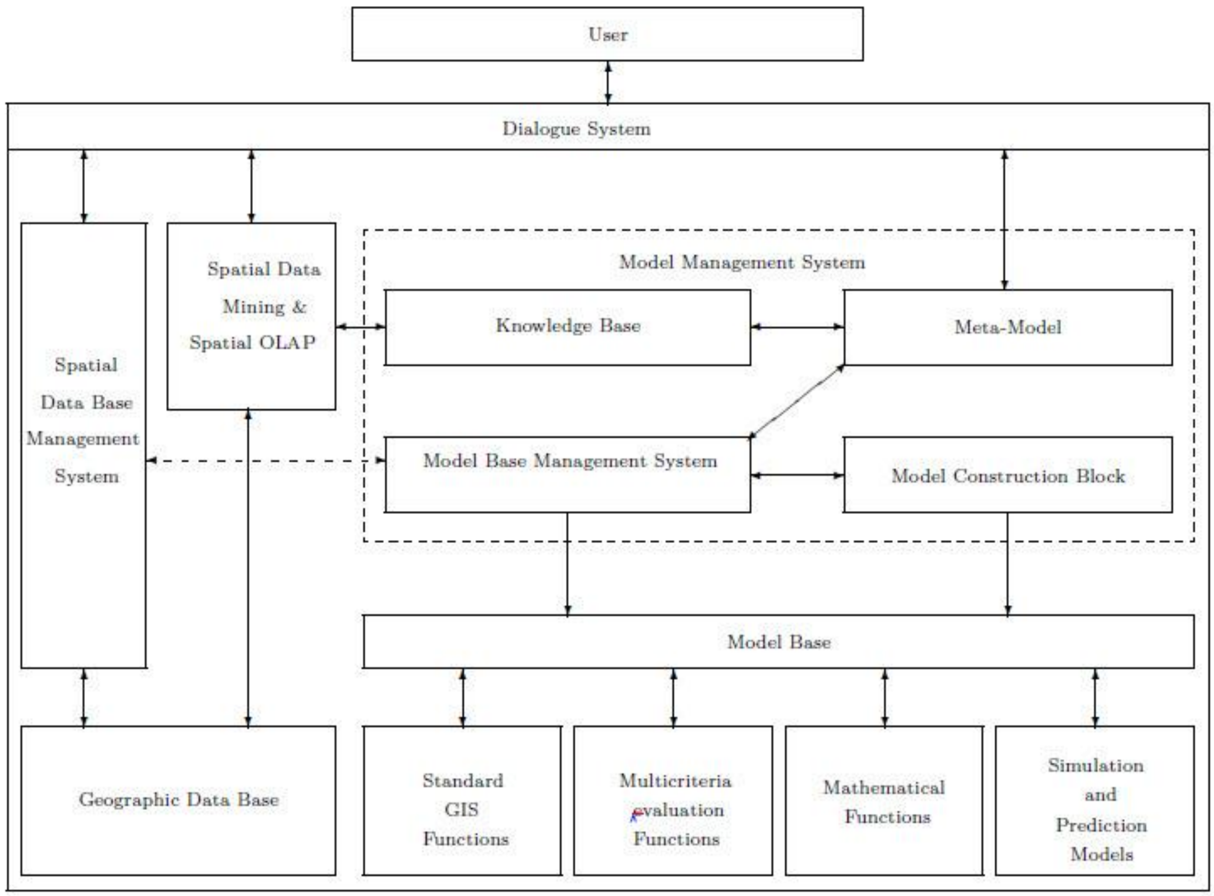


Fig 2.7: A design of a multicriteria SDSS [Chakhar & Mousseau, 07]

[Chakhar & Mousseau, 09] proposed an approach to generate a number of alternatives based on a planar subdivision of the study area. The planar subdivision is obtained by combining a set of criteria maps. The result is a set of non-overlapping polygons/spatial units. Punctual, linear and areal decision alternatives, conventionally used in **spatial multicriteria analysis**, are then constructed as an individual, a collection of linearly adjacent or a collection of contiguous spatial units. This alows to reduce substantially the number of alternatives enabling the use of outranking methods.

[MARINONI, 05] proposed A stochastic spatial decision support system based on PROMETHEE. PROMETHEE is a well-known multicriteria decision support method which gives insight into the preference structure of a whole set of alternatives. The author discussed the integration of PROMETHEE in a GIS. He have shown how the combination of a sophisticated decision support methodology with spatial analysis and visualization capabilities can be applied to evaluating decision alternatives that are made up of regular or irregular shaped zones. Moreover, the author presented a framework which facilitates the application of the proposed approaches.

4 Intelligent Decision Support System

Intelligent environmental decision support system (IEDSS) *"is an intelligent information system that reduces the time in which decisions are made in an environmental domain, and improves the consistency and quality of those decisions"* [HAAGSMA & JOHANNS, 94]

"The use of Artificial Intelligence tools and models provides direct access to expertise, and their flexibility makes them capable of supporting learning and decision making processes. There integration with numerical and/or statistical models in a single system provides higher accuracy, reliability and utility" [CORTÉS ET AL., 00],
So IDSS = DSS + AI.

4.1 Artificial intelligence

According to the authors Artificial Intelligence seems the thing that make the difference. John McCarthy [MCCARTHY, 07] defined it as*" It is the science and engineering* *of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable"*. He coined the term in 1955, [McCarthy et al, 55] defined it as *"the science and engineering of making intelligent machines"*

And about **intelligence** definition *"Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines"*. [McCarthy, 07]

4.2 DSSs categorization based on degree of intelligence

DSSs can be categorized based on the degree of intelligence provided as shown in Table 1.2

Category	Domain	Example
Data	Data driven	mining Lists, data bases and data management systems.
Evolutionary	Genetic algorithms	Optimization and search spaces
Thought	Neural networks	Learning and pattern matching
Constraints	Rule based systems	Expert and knowledge based systems
Symbolic	Fuzzy logic	Transforming the ambiguity into fuzzy sets
Temporal	Case based systems	Reasoning and analogy based systems
Iterative	Inductive Machine learning	creation of dynamic rule sets

Tab 2.2 Decision Support System domains categorized [Dhar & Stein, 1997] [TWEEDALE & AL, 2016]

Before moving to the definitions of domains, we talk about a huge domain that can use and combines other many fields of the previous domains (Tab 2.2) and techniques which is **data mining**.

4.3 Data mining

Data mining is defined as *"the process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities."* [**WITTEN & FRANK, 05**]

Data mining is about solving problems by analyzing data already present in databases.

D.J.Hand [**SUMATHI & SIVANANDAM, 06**] said about Data mining: *"I shall define Data Mining as the discovery of interesting, unexpected, or valuable structures in large data sets"*

In the definition, the author use the adjectives *interesting, unexpected* and *valuable* to describe the discovered structure. When we talk about unexpected things we almost mean the future. Data Mining aims to explain past experiences and also make predictions to the future by data analysis. The combination of many domains (**statistics**, **machine learning**, **artificial intelligence** and **database technology**) gives Data mining the character of multi-disciplinary.

Nowadays Data mining applications are extremely valuable. over years of work the majority of companies have stored voluminous amount of data, Data mining can extract very valuable structure (knowledge) from this data which was neglected (explain the past).

Then the companies are able to control the extracted knowledge to have more clients, more sales, and greater profits by making better decisions. Data mining represent also very powerful tool in engineering and medical fields.

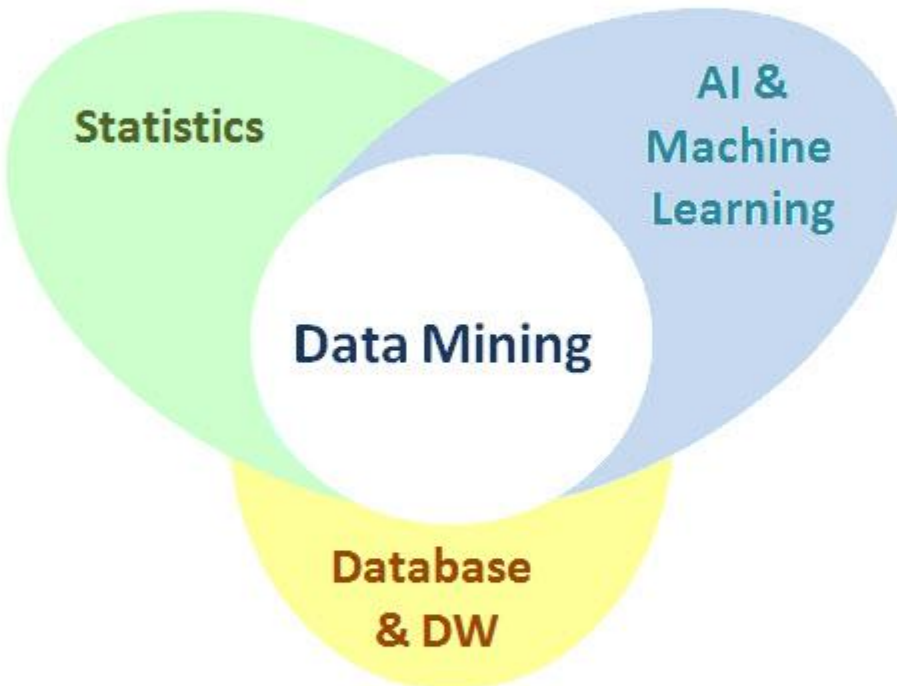


Fig 2.8: Data mining as a multi-disciplinary field [Sayad, 2016]

Statistics

The science of collecting, classifying, summarizing, organizing, analyzing, and interpreting data. [MERRILL, 2012]

Artificial Intelligence

We have already explained this concept.

Machine Learning

We will explain this concept with details in the next section.

Database

We don't mean by Database the theoretical definition: computer science container which permits you to store your data (collection of organized data) … but the science and technology that allows users update their data (add, modify, remove) in addition to collect, store and manage it.

Data warehousing

Can be defined as the responsible process for the design, administration, construction and refresh the data warehouse, with advanced multi-dimensional reporting that support decision making processes. (See [INMON, 05])

In the academic domain and research domain Knowledge Data Discovery KDD is different of Data mining but in industry domain which represented in the big firms (Microsoft, Google, Facebook…) the two concepts are similar.

KDD *"is the nontrivial process of identifying valid, novel, potentially useful, and ultimate understandable patterns in data"* [FAYYAD & AL, 96]

As we have mentioned about Data mining in research domain, [FAYYAD & AL, 96] confirm that Data mining considered also as a step in the KDD process that consists of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns (or models) over the data. Note that the space of patterns is often infinite, and the enumeration of patterns involves some form of search in this space. Practical computational constraints place severe limits on the subspace that can be explored by a data-mining algorithm. [FAYYAD & AL, 96]

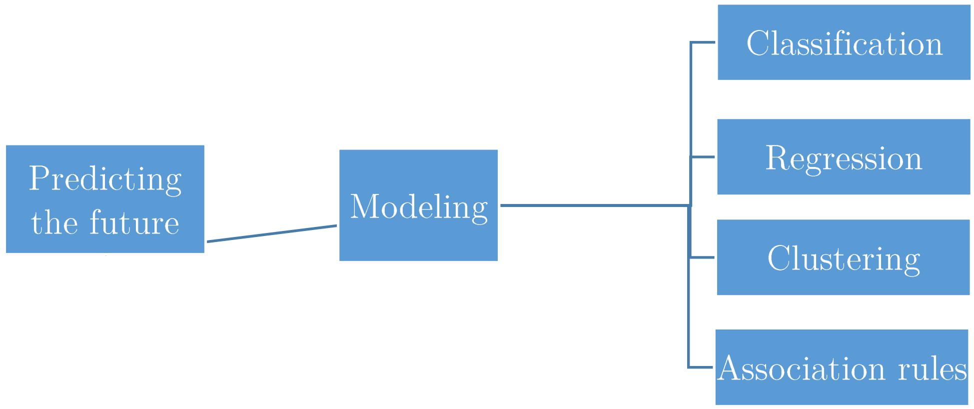


Fig 2.9 data mining tasks

4.3.1 Machine Learning

[**Mitchell, 97**] consider machine learning as a multidisciplinary field, that draws on results from research fields (mainly Data mining) as diverse as Artificial Intelligence, Bayesian methods, Computational complexity theory, Control theory, Information theory, Philosophy, Psychology, Neurobiology [**Mitchell, 97**]

Machine learning that allows a machine to learn to perform tasks from a learning base containing examples already treated. Each element of the training set (training set) is an input-output torque.

There is two types of learning in data mining: **supervised learning** that is used primarily for data classification and **unsupervised learning** which is used in the search for associations or groups of individuals.

Supervised learning (Classification learning) primarily relates to data classification methods (we know the input and the output is to be determined) and regression (you know the output and we want to find the entrance). [**WITTEN & FRANK, 05**]

The primary machine learning approaches are Decision Trees, Artificial Neural Networks, Learning set of rules (Expert systems), Inductive Logic Programming, Instance-Based Learning, similarity measure and Genetic Algorithms…

Supervised learning: Decision trees, neural networks.

Unsupervised learning: Clustering, Associative rules Sequence mining.

[TWEEDALE & AL, 2016]

4.3.1.1 Related work

[YANG ET AL, 2015] shows a new approach in order evaluate potential sites for suggested hotel properties by designing an automated web GIs application: Hotel Location Selection and Analyzing Toolset (HoLSAT) .a set of machine learning algorithms is used predict several business success indicators associated with location sites. Using an example of hotel location assessment in Beijing, HoLSAT calculates and visualizes various desirable sites contingent on the specified characteristics of the proposed hotel. Thus, the suggested approach proves its use fullness in the field of hotel location evaluation.

4.4 DSSs domains

After having an idea about data mining, moving now to the domains Tab 2.2.

4.4.1 Data-Driven

Data-driven methodologies are **data mining** techniques based methodologies. The inputs need to be associated, organized and queried previously to being interpreted, they are based on a list of related facts. For example, On-line Analytical Processing (OLAP) in Data warehouse, Executive Information Systems (EIS) and Geographic Information Systems (GIS) are types of data-driven DSS. Model-driven spatial DSS are also common **[POWER, 07]**

RELATED WORK

[KISILEVICH ET AL, 2013] present a GIS-based decision support system that can both, estimate objective hotel room rates using essential hotel and locational characteristics and predict temporal room rate prices. For an objective comparison, information about objective hotel room rates provides the basis for a realistic computation of the contract's profitability .this latter can help monitoring past hotel room rates and adjusting the price of the future contract.This research makes three major contributions [KISILEVICH ET AL, 2013].

[ROCKART, 79] stimulated the development of executive information systems (EIS) and executive support systems (ESS).both systems are evolved from unique user model. Driven decision support systems and from the evolvement of relational data base products. Analysts for senior executives used the first EIS predefined information screens maintained. For example, in the fall of 1978, development of an EIS called Management Information and Decision Support (MIDS) system began at Lockheed-Georgia [HOUDESHEL & WATSON, 1987]

4.4.2 Genetic Algorithms

Data processing is originated as a result of an excessive number of parameters. A new approach is called to be used by the exponential growth of combinatorial statistics .The use of *Darwinian Theory* relating to *survival of the fittest* concept was proven to efficiently find the best solution of a given probable subset and has been adopted to for many search space problems. [TWEEDALE ET AL, 2016]

4.4.3 Decision trees

A decision tree is a graphical representation without oriented cycle. The internal nodes of the tree are tests on the fields or attributes, the leaves are classes and arcs are the classes of the source class partitioning predicates.

The classes are inferred from the data. They can be defined as a logical rules. The order of attributes plays a crucial role in the construction of the tree. At each change of order, a new tree is generated.

Decision tree learning is one of the most widely used and practical methods for inductive inference. It is a method for approximation of discrete-valued functions, in which a tree represents the learned function.

Decision trees algorithms CART, ID3, C4.5, OC1, SLIQ, SPRINT …

4.4.4 Neural Networks

They are based on the principle of DNA where genes combine together and weakest disappear (principle of natural selection: selection, reproduction and mutation).

A neural network simulates cognitive approach to learning using trial and error. It involves the creator maintaining a prediction loop, which can use comparisons and an adjustment to process to reconcile transfer functions. Forward and backward chaining neural networks have been created to minimize overall network errors. **[TWEEDALE & AL, 2016]**

Artificial neural networks algorithms: AdaBoost, Learn++ …

Related work

In **[YOUNSI & AL, 07]** the authors propose decision model for territory management by using GIS and Artificial Neural Networks (ANN). This model has being accompanied by business intelligence approach.

4.4.5 Rule Based Systems

[TWEEDALE & AL, 2016] say that rule based Systems (like *Internist* and *XCON)*, use a combination of knowledge management tools (ultimately termed KBS), working memory and rule interpreters to distinguish between a collection of facts represented as evidence in knowledge bases and queries input by operators. Forward chaining (branching down the tree) and backward chaining (evidence matching up through working memory). [TWEEDALE & AL, 2016]

4.4.6 Fuzzy Logic

Nowadays, the existing systems fail to approximate information the way humans interpret then surroundings (especially RBS). Fuzzy logic introduces natural language membership functions, based on underlying truth statements represented by the Boolean values of '0' or '1'.For instances , we can express a numerical temperature reading as gradations between "hot" and "cold". This scale could also be expanded to include 'warm' or 'cool'. Therefore, humans are enabled to accept information (measure against a range, without being burdened with the need to process minor errors or fluctuations). The membership is transformed from *crisp sets* into *fuzzy sets*. [TWEEDALE ET AL, 2015]

RELATED WORK

[WAN ET AL, 90] describe a method, a representation of information and fuzzy transformation in GIS context that offer the possibility of elaborate a fuzzy evaluation method.

4.4.7 Case Based Reasoning

A Case Based Reasoning (CBR) system takes advantage of previous attempts to solve a problem or improve the accuracy of the system. Training is gained using a predefined data set and matured using new patterns as time passes. Reasoning is generated using a series of measures that are used to describe the situation within a scenario. Cases are determined using a pattern matching approach of analysis. [TWEEDALE ET AL, 2016] (Advisory Systems)

4.4.8 Naïve Bayes

Bayesian classifiers are statistical classifiers. Which can predict class membership probabilities, such as the probability that a given sample can be a member a particular class. Bayesian classifier is based on Bayes' theorem. Naïve Bayesian classifiers are based on the idea of independency. i.e, the effect of an attribute value on a particular class is independent of the valves of the other attributes. **[LEUNG¸07]**.

Related work

[PHIT ET AL, 2010] introduce a predictive model for the number of vacant taxis in a specific area based on time of the day, day of the week, and weather condition. The history is used to build the prior probability distributions for our inference engine, which is based on the naïve Bayesian classifier with developed error-based learning algorithm and method for detecting adequacy of historical data using mutual information. Based on 150 taxis in Lisbon, Portugal, **[PHIT ET AL, 2010]** able to predict for each hour with the overall error rate of 0.8 taxis per 1x1 km2 area.

[PARK & STENSTROM, 08] investigated Bayesian networks to classify urban land use from satellite imagery. They used land sat enhanced thematic Mapper plus (ETM) images for the classification in two study areas. Bayesian networks provided 80–95% classification accuracy for urban land use using four different classification systems. The classifications were robust with small training data sets with normal and reduced radiometric resolution. The network was able of utilizing information from non-spectral data and the relation among variables is explicitly shown. We can use this classification to provide timely and inexpensive land use information over large areas for environmental ains such as estimating storm water pollutant loads.

4.5 Other related works (combination of many techniques)

Other authors combine two techniques or more for their works. [GUANLIN ET AL, 2015] Propose an IAMSULI (Intelligent Analysis and Mining System for Urban Lighting Information). Using data analysis technology, case-based reasoning technology and data mining technology (decision trees) comprehensively, the system aims at providing better decision support for lighting management. The proposed system [GUANLIN ET AL, 2015] is composed of system management, geographical information management, lighting information management, lighting data analysis, comprehensive classification statistics and event management. It can help quickly to find lighting problems, provide early warning alarm of the problems, and also supply constructive information for the urban development.

[GUILHERME & GILBERTO, 2013] presents a supervisor agent for urban traffic control, which includes fuzzy sets, genetic algorithms and case-based reasoning to create new programs for traffic controllers. The results showed that the strategy used may be adaptive and help improve traffic conditions by diminishing cars queues in cities' intersections.

[GOLDING & ROSENBLOOM, 91] propos an architecture for combining rule-based and case-based reasoning. It aims to improving the performance of rule-based systems through case-based reasoning (CBR).

[KAR, 2015] presents the application of a hybrid approach for group decision support for the supplier selection problem. The author integrated Fuzzy set theory, analytic hierarchy process (AHP) multi-criteria decision method and neural networks to provide group decision support under consensus achievement. [KAR, 2015] also use Discriminant analysis for the purpose of supplier base rationalization, through which suppliers have been mapped to highly suitable and less suitable supplier classes. The proposed integrated approach has been further studied through two case studies and the proposed approach has been compared with another approach for group decision making under consensus and other approaches for prioritization using AHP, without consensus achievement. [**KAR, 2015**] affirm that very high accuracy in capturing the collective consensual preferences of the group was established across eight cross-validation tests from the two case studies, for the hybrid approach, even with extremely limited count of data sets which were used for training the hybrid model.

[**ASADI, 2015**] develops a new rule-based decision support system (RB-DSS) to find the safest solutions for routing, scheduling, and assignment in Hazmat transportation management.

The author's system (RB-DSS) aims to define the safe program, the accident frequency and severity are estimated for different scenarios of transportation, and they are used to classify the scenarios by a new structure of decision tree (DT), which is proposed to select branching variables at the primary levels according to the experts' perception. The outputs of the DT are stated in the form of if–then rules trained by a multilayer perceptron neural network to generalize the safe programs for Hazmat transportation.

[**CALZADA ET AL, 2015**] utilizes a recently-developed spatial intelligent decision system (IDS) based on rule based system and GIS, the IDS named, Spatial RIMER$_+$, to model the self-rated health estimation decision problem using real data in the areas of Northern Ireland, UK. The goal is to learn from past or partial observations on self-rated health status to predict its future or neighborhood behavior and reference it in the map. Three scenarios in line of this goal are discussed in details, i.e., estimation of unknown, down scaling, and predictions over time. They are used to demonstrate the flexibility and applicability of the spatial decision support system and their positive capabilities in terms of accuracy, efficiency and visualization.

4.6 Integration of Big Data in DSSs

4.6.1 What is big data?

According **[O'LEARY, 2013]**Michael Cox and David Ellsworth1 were among the first to use the term *big data* literally, referring to using larger volumes of scientific data for visualization (the term *large data* also has been used). Currently, there are a number of definitions of big data. Perhaps the most well-known version comes from IBM, which suggested that big data could be characterized by any or all of three "V" words to investigate situations, events, and so on: volume, variety, and velocity. **[O'LEARY, 2013]**

Volume refers to larger amounts of data being generated from a range of sources. For example, big data can include data gathered from the Internet of Things (IoT). As originally conceived, IoT referred to the data gathered from a range of devices and sensors networked together, over the Internet. Big data can also refer to the exploding information available on social media such as Facebook and Twitter. We must be able to treat them!

Variety refers to using multiple kinds of data to analyze a situation or event. On the IoT, millions of devices generating a constant flow of data results in not only a large volume of data but different types of data characteristic of different situations.

Sources, shapes and very different formats, structured and unstructured: Also referred to as complex data. **[PINGDOM, 2011]** so data must be treated simultaneously!

Velocity of data also is increasing rapidly over time for both structured and unstructured data, and there's a need for more frequent decision making about that data. As the world becomes more global and developed, and as the IoT builds, there's an increasing frequency of data capture and decision making about those "things" as they move through the world. Further, the velocity of social media use is increasing. For example, there are more than 250 million tweets per day.4 Tweets lead to decisions about other Tweets, escalating the velocity. Further, unlike classic data warehouses that generally "store" data, big data is more dynamic. As decisions are made using big data, those decisions ultimately can influence the next data that's gathered and analyzed, adding another dimension to velocity. Big data isn't just volume, variety, and velocity, though; it's volume, variety, and velocity at *scale*. As a result, big data has received substantial attention as distributed and parallel computing has allowed processing of larger volumes of data, most notably through applications of Google's Map Reduce.

Frequent updates, incoming data stream, and rapid obsolescence of certain data ... need analysis in near real time (eg. Detection / prevention failures, queue management).

We must treat them quickly!

4.6.1.1 Related work

[LI ET AL, 2012] proposes an improved ARIMA based prediction method to forecast the spatial-temporal variation of passengers in a hotspot. Evaluation with a **large-scale** real-world **dataset** of 4 000 taxis' GPS traces over one year shows a prediction error of only 5.8%. **[LI ET AL, 2012]** also explore the application of the prediction approach to help drivers find their next passengers. The simulation results using historical real-world data demonstrate that, with the guidance, drivers can reduce the time taken and distance travelled, to find their next passenger, by 37.1% and 6.4%, respectively.

[OFLI ET AL, 2016] proposed a hybrid crowdsourcing and real-time **machine learning** solution to rapidly process **large volumes** of aerial data for disaster response in a time-sensitive manner. Crowdsourcing can be used to annotate features of interest in aerial images (such as damaged shelters and roads blocked by debris). These human-annotated features can then be used to train a supervised machine learning system to learn to recognize such features in new unseen images. **[OFLI ET AL, 2016]** described how this hybrid solution for image analysis can be implemented as a module (i.e., Aerial Clicker) to extend an existing platform called Artificial Intelligence for Disaster Response (AIDR), which has already been deployed to classify microblog messages during disasters using its Text Clicker module and in response to Cyclone Pam, a category 5 cyclone that devastated Vanuatu in March 2015.

The results suggest that the developed platform which combine crowdsourcing and **machine learning** to make sense of **large volumes** of aerial images can be used for disaster response.

5 Discussion

The related works mentioned above (dispatched) have advantages as they have drawbacks or limitations, whether those of the methods of multiple-criteria decision analysis, those which integrated data mining techniques or those which combines several techniques and integrated the Big Data.

5.1 Using of MCDA methods

The use of MCDA methods can lead to many possible disadvantages. An incredible amount of input is necessary at every step of the procedure in order to accurately record the decision maker's preferences (**Maut**). For example the problem of ELECTRE is that its process and outcomes can be hard to explain in layman's terms. Further, due to the way preferences are incorporated, the lowest performances under certain criteria are not displayed. The outranking method causes the strengths and weaknesses of the alternatives to not be directly identified, nor results and impacts to be verified [**KONIDARI & MAVRAKIS, 07**]. The PROMETHEE's problem is that it does not provide a clear method by which to assign weights and it requires the assignment of values but does not provide a clear method by which to assign those values. According to [**VELASQUEZ & HESTER, 2013**] AHP

Problems due to interdependence between criteria and alternatives; can lead to inconsistencies between judgment and ranking criteria; rank reversal. In the other hand the problem of expansion of comparisons when we have a large number of criteria.

5.2 Using Multi Agent Systems (MAS)

The use of Multi Agent System responds the concept of intelligent in system and considered as very powerful tool, because they are (agents) autonomous, but they have also many drawbacks. One of the main weakness of M.A.S is the heavy load caused by communication between the agents especially when we talk about huge volume data, then the amount of communication make MAS out of control.

5.3 Using Data mining techniques

Data mining represent a very powerful and performing tool that can be used in decision support as showing in the related works but also have some disadvantages depends to the used techniques. Case based reasoning (CBR) for example is sensitive to inconsistent data; requires many cases. Fuzzy set theory is difficult to develop; can require numerous simulations before use [**VELASQUEZ & HESTER, 2013**]. So each method have a points of weakness as it have advantages.

[**AKINOLA ET AL, 2015**] studied the performances of machine learning methods (algorithms): Decision Tree, Multi-Layer Perceptron and Naïve Bayes classification with respect to their times taken for training and accuracy of prediction. The study shows that even though Naïve Bayesian algorithm takes less time for its prediction although the huge amount of data used in the experimentation. Naïve Bayesian models are popular in machine learning applications, due to their simplicity in allowing each attribute to contribute towards the final decision equally and independently from the other attributes. This simplicity equates to computational efficiency, which makes Naïve Bayesian techniques attractive and suitable for many domains [DANIELA & AL, 09].

5.4 Synthesis

Work	Family	Technique/Method	Drawbacks
[CHAKHAR & MARTEL, 03] [MARINONI, 05] [CHAKHAR & MOUSSEAU, 07]	MCDA	ELECTRE	Process and outcomes can be hard to explain in layman's terms. [KONIDARI & MAVRAKIS, 07]
		AHP	Expansion of comparisons due the number of criteria
		PROMETHEE	indifference thresholds has not concrete interpretation to the decision maker [KONIDARI & MAVRAKIS, 07]
[GUILHERME & GILBERTO, 2013]	Multi Agent System	Combining MAS and other techniques	heavy load caused by communication between the agents
[YOUNSI ET AL, 07] [WANG AND WU, 09] [PHIT ET AL, 2010] [PARK & STENSTROM, 08]	Data Mining	Case based reasoning (CBR) And Rule based (RBS)	Sensitive to inconsistent data; requires many cases.

		Artificial Neural Network (ANN)	black box that does not explain its decisions
		Fuzzy logic	difficult to develop; can require numerous simulations before use
		Genetic Algorithm	are expensive in time calculation as they handle multiple solutions simultaneously
[KAR, 2015] [ASADI, 2015] [CALZADA ET AL, 2015]	Combination of many techniques	Data mining techniques, MCDA techniques and MAS	
[LI ET AL, 2012] [OFLI ET AL, 2016]		Integration of Big data	

Tab 2.3 Classification of the related works

Our project aims to offer an architecture and intelligent system that can learn from previous uses to increase the performance of decision support, dedicated to the planning staff (Decision makers).we will try to offer better performance of decision support, not with the MCDA methods but using other techniques. **[SPATZ ET AL, 2014]** said that intelligent system is a system that can perceive, create action, and **learn** in an autonomous fashion, i.e., without external supervisory intervention for an extended amount of time. And **[YANG ET AL, 2012]** said that Data mining increases the "intelligence" of DSS. Without doubt at the end of this discussion you have almost an idea of what could we use in contribution.

6 Conclusion

We have talked about several topics in this chapter which was mainly the decision support especially spatial domain, DSSs and IDSS with their different techniques intelligent agent, data mining methods like machine learning etc… Also related works for each domain of them.in the next chapter we will present our contribution and the different used techniques.

Chapter 3

Proposed architecture of IDSS for urban planning

1 Introduction

The aim of this chapter is to detail our contribution in this thesis, which involves the design and the implementation of IDSS (decision model or architecture, and associated process) which uses data mining as a way to bring out the intelligence dedicated to urban planners. Starting with the architecture, moving to the components.

2 Goals

In the context of the general problem that concerns us, we must respond in a primary objective and specific objectives, respectively:

2.1 Primary objective

The main objective is to propose an architecture and develop an Intelligent Decision Support which can help decision makers (urban planning staff) by organize and synthesize such information takes into account the opinion of citizens in a way which leads them (Decision makers) to feel comfortable and confident about making a decision.

The system can learn, and perceive in an **autonomous fashion**, i.e., without external supervisory intervention and change his behavior from use to another. He learn from previous uses to increase the performance of decision support, dedicated to the Decision makers. We will try to offer better performance of decision support, not with the MCDA methods but using other techniques.

2.2 Specific objectives:

- Try to exit the dependency to MCDA methods in solving decision problems.
- Use high level techniques to solve problems, such as data mining techniques: machine learning → naïve Bayes classifier, clustering…

- Use high performance distributed data base dedicated to large volume: HBase.
- Learn from large data set.

After fixing our goals, both primary and specific. The next section will contains our proposed architecture for the Intelligent Decision Support System with description for each component of it.

3 System Architecture

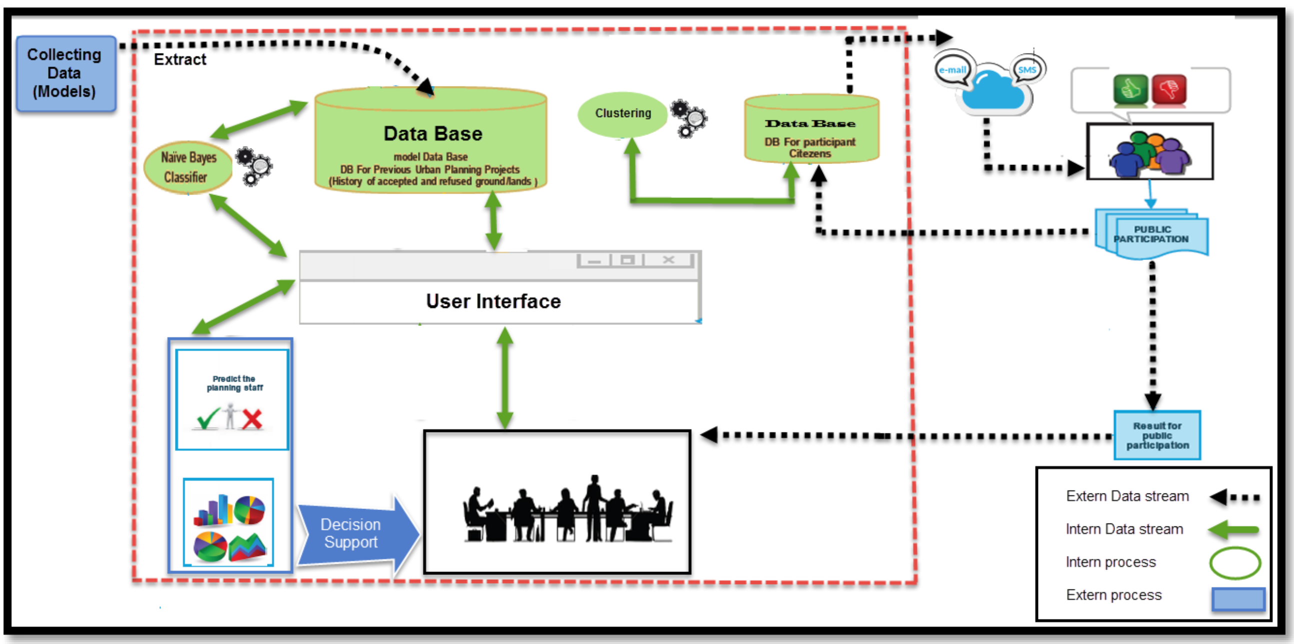


Fig 3.1 Proposed architecture

3.1 Description

To ensure best comprehension to the architecture we'll bring out the sequencing of actions that does not appear (it's normal because it's not a process). The user will first enter the alternatives (lands/fields and its characteristics in our study case) through the user interface. Alternatives will be copied in the database (model base). An internal process is launched which is a classification based on naive Bayes. This process has as result a prediction of the consequences whether they will be good or bad (accepted / refused in our study case), in addition to their probabilities presented as percentage. The results are based on previous experiments that exists in the model base and which are collected previously (extern process: collecting data (models)). The results will be shown and displayed to the user graphically with bar chart and pie chart and other presentation, aiming with this to support the user to make a decision. (The user will be asked to confirm his choice) through the GUI, user can select his choice among the alternatives and enter the citizen satisfaction (we'll see how). The alternatives that are already copied previously in the model base (are not regarded as models in the previous classification) will be validated, the user choice as accepted and the other as denied. So now the land have become models and they will be used for improving the results (the system will perceive and learn). If you reenter the same data (fields with same characteristics) the system will react differently because of new models.

That what concerns the first part, to the other part that both internal and external phase (the limits of the system are the lines in red). Our idea from the beginning includes the participation of citizens in the planning process. Citizens are supposed to give their opinions through SMS or Email or vote on a website is that they are satisfied or not. The result here is: the opinion of every citizen. We propose to manage the database of citizens or can be a clustering for them such that we detect(obtain) different categories: active, passive… we get the citizens who are not satisfied and try to know their desires and tendencies and sure take it on consideration, this process concerns citizens who vote more than the threshold defines. In the picture (Fig 3.1) there is three extern process: public participation, result for public participation and collecting data (models). The difference between public participation and result for public participation is that the first is the opinion of each participant, and the second is result of the first, in other term is the opinion of the all citizens, for example the participant (citizens) are satisfied for this project. This will be entering the validation phase. The other extern process collecting data (models) is process in which we collect (get it if exist) a history of the fields, whether accepted or rejected, which were previously selected as alternatives for projects.

The outcomes will be used to improve the performance, in the too principle internal processes Naïve Bayes and the clustering. So he can perceive, learn and change his behavior autonomously.

4 Architecture components

In this section we presents the different components of our architecture.

4.1 Database

Data Base in our architecture intuitively is used to store the data (urban in our case). In our case the data stored as models (to be learned later), so in such way is a model base. The nature of crucial urban data (required data for planning) is voluminous and massive and may be heterogenic and unstructured (later), millions of streets, millions if we don't say billions of lands with different types that already used and others not yet, millions of underground Installation varies between underground cables (electricity, phones…) and sanitation other installations millions of projects under construction. In front to all this, we have to think about a tool that have the ability to store this huge amount of data. Without a doubt it will be a Big Data tool (we will see this in the next chapter…).

4.2 Naïve Bayes classifier

Starting with the first componnent which apear in the architecture (Fig 3.1) as a intern process. It represent mainly our contribution naïve bayes classifier

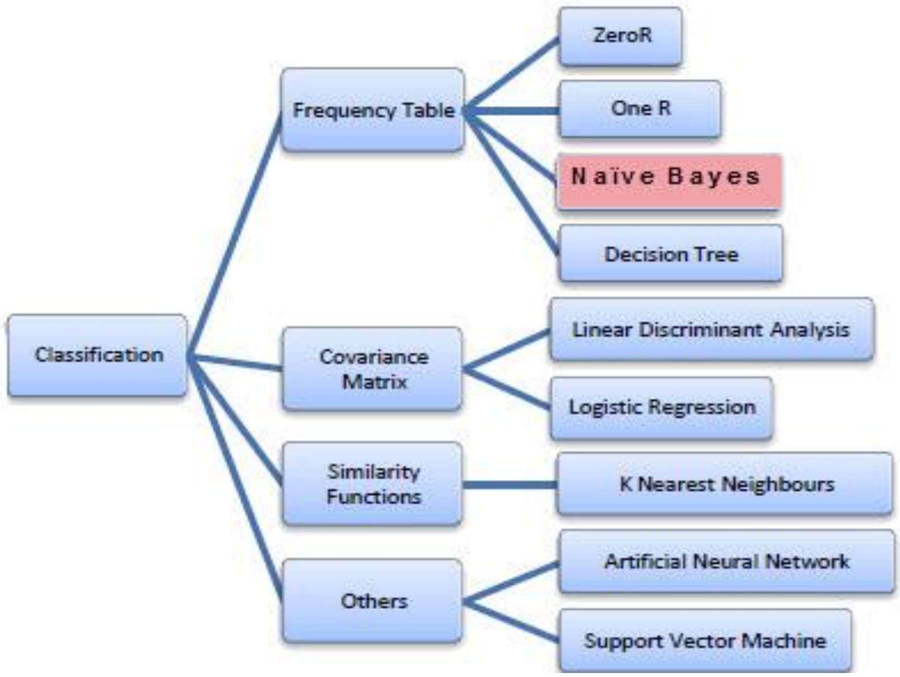


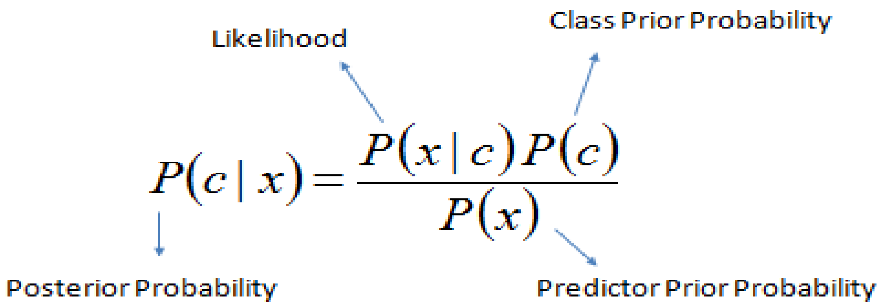
Fig 3.2 Data mining classification types [Sayad, 2016]

We did not speak a lot of this classifier in the state of the art (chapter 1), we have to give him what he deserves from the discussion. Some specialists [LEUNG¸07], [Sayad, 2016] have defined this classifier, let's have a look to what they say about it

"Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given sample belongs to a particular class. Bayesian classifier is based on Bayes' theorem. Naive Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computation involved and, in this sense, is considered "naive". "[LEUNG¸07]

4.2.1 Bayes' Theorem

Bayes theorem provides a way of calculating the posterior probability, $P(c/x)$, from $P(c)$, $P(x)$, and $P(x/c)$. Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class C (*C1, C2, C3...Cn*) is independent of the values of other predictors. This assumption is what we called it class conditional independence.



$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Fig 3.3 Bayes theorem

4.2.2 Demarche

For more comprehension of the theorem, let's have a look to this example that describes the lands (fields) conditions for building a project. The goal of this example is to get the posterior probability. The posterior probability can be calculated by first, constructing a frequency table for each attribute against the target. Then, transforming the frequency tables to likelihood tables and finally use the Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

NB: the training data (Tab 3.1) set represent a sample from our data set (just four attributes).

Terrain Nature	**Existing pole**	**Juridical nature**	**Underground cables**	**Decision**
Agricultural	High voltage	Private	Non-existent	**Refused**
Agricultural	High voltage	Private	Exist	**Refused**
Buildable	High voltage	Private	Non-existent	**Accepted**
Clayey	Medium voltage	Private	Non-existent	**Accepted**
Clayey	Low voltage	Domaniale	Non-existent	**Accepted**
Clayey	Low voltage	Domaniale	Exist	**Refused**
Buildable	Low voltage	Domaniale	Exist	**Accepted**
Agricultural	Medium voltage	Private	Non-existent	**Accepted**
Agricultural	Low voltage	Domaniale	Non-existent	**Accepted**
Clayey	Medium voltage	Domaniale	Non-existent	**Refused**
Agricultural	Medium voltage	Domaniale	Exist	**Accepted**
Buildable	Medium voltage	Private	Exist	**Accepted**
Buildable	High voltage	Domaniale	Non-existent	**Accepted**
Clayey	Medium voltage	Private	Exist	**Refused**
Buildable	Low voltage	Private	Non-existent	**Accepted**
Buildable	Low voltage	Domaniale	Exist	**Refused**

Tab 3.1 Sample of the data set

Nature Terrain		**Existing pole**		**Juridical nature**		**Underground cables**	
Accepted	**Refused**	**Accepted**	**Refused**	**Accepted**	**Refused**	**Accepted**	**Refused**
Agricultural 2	4	High 2	5	Private 4	6	Inexist 8	4
Buildable 6	2	Medium 5	2	Domaniale 7	3	Exist 2	6
Clayey 3	3	Low 4	2				

Decision	
Accepted	Refused
11	9

Tab 3.2 the frequency table

Terrain Nature		**Existing pole**		**Juridical nature**		**Underground cables**	
Accepted	Refused	Accepted	Refused	Accepted	Refused	Accepted	Refused
Agricultural 2/11	4/9	High 2/11	5/9	Private 4/11	6/9	Inexist 8/11	4/9
Buildable 6/11	2/9	Medium 5/11	2/9	Domaniale 7/11	3/9	Exist 2/11	6/9
Clayey 3/11	3/9	Low 4/11	2/9				

Decision	
Accepted	Refused
11/20	9/20

Tab 3.3 Likelihood Table

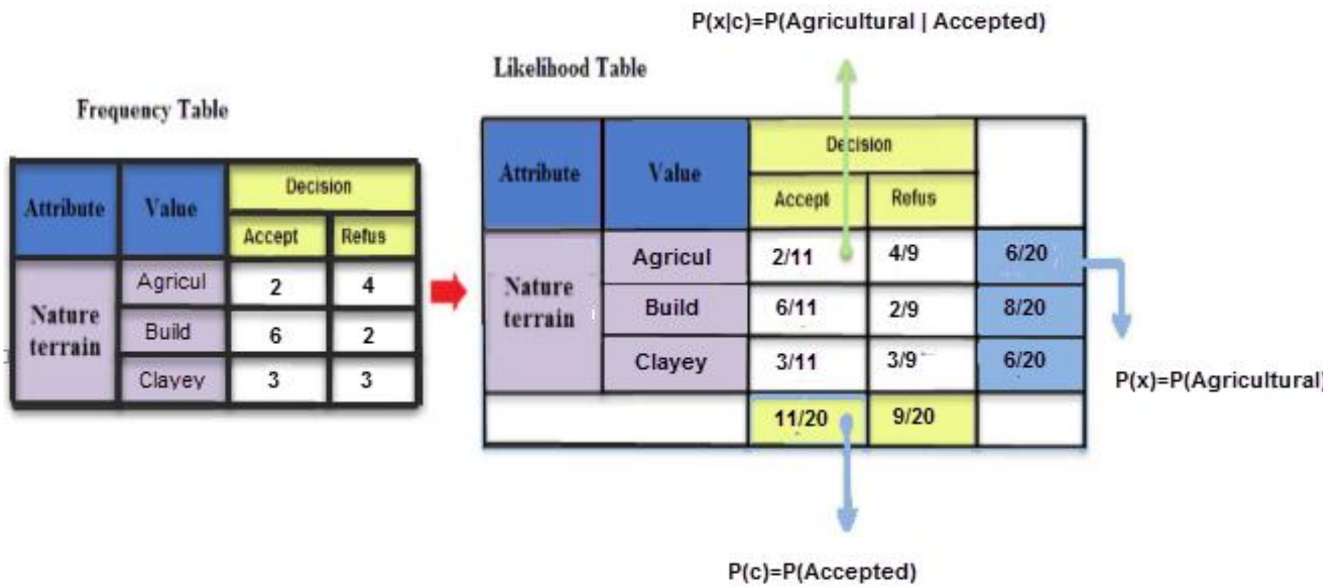


Fig 3.4 Calculate posterior probability for a single attribute

Now assume that we have to classify the following new instance:

TERRAIN NATURE	EXISTING POLE	JURIDICAL NATURE	UNDERGROUND CABLES	DECISION
AGRICULTURAL	MEDIUM VOLTAGE	PRIVATE	EXIST	?

Tab 3.4 New instance

Based on the probability distribution in the training data, we calculate a probability for each class. First take into account the probability of each attribute.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$Theorem: P(c|x) = P(x_1|c) \times P(x_2|c) \times P(x_3|c) \dots P(x_n|c) \times P(c)$$

P (Accepted) = 2 /11 × 5/11 × 4/11 × 2/11 × 11/20 = 0.00546

P (Refused) = 4/9 × 2/9 × 6/9 × 6/9 × 9/20 = 0.049

Now we normalize:

P (Accepted) = 0.003/ (0.003+0.022) = 0.12 →12%

P (Refused) = 0.022/ (0.003+0.022) = 0.88 →88%

Now if we want to classify this instance, we choose the class so that it maximizes this probability. This means that we predict that the decision of new instance will be **Refused**.

Offer the decision maker the decision support and make him more confident when he make a decision is what we try to do.

The classification above, seems good for a single instance, but how can we choose between multiple instances (fields in our project).

The idea is to calculate the posterior probability for each instances and take the values of the class accepted “Accepted” (1). After that we normalize them so that all the values or the percentage of such that the sum of percentages is 100% or 1. Here is the steps how it work.

TERRAIN NATURE	EXISTING POLE	JURIDICAL NATURE	UNDERGROUND CABLES	DECISION
BUILDABLE	HIGH VOLTAGE	DOMANIALE	EXIST	?
BUILDABLE	LOW VOLTAGE	PRIVATE	INEXISTANTE	?
BUILDABLE	MEDIUM VOLTAGE	DOMANIALE	INEXISTANTE	?

Tab 3.5 New instances (alternatives)

1/ Calculate probability for class "Accepted"

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times P(x_3|c) \dots P(x_n|c) \times P(c)$$

P1 (Accepted) = 0.0063

P2 (Accepted) = 0.0289

P3 (Accepted) = 0.0631

2/Calculate sum

Sum = $\sum_1^n \boldsymbol{P_i}$ / **i:** is the instance index

Sum = 0.0983

3/Calculate percentage for each instance

$$\text{Percentage(i)} = \frac{P_i}{Sum}$$

Percentage (1) = 0.0063 / 0.0983= 0.064 → 6.4 %

Percentage (2) = 0.0289 / 0.0983= 0.2939 → 29.4 %

Percentage (3) = 0.0631/ 0.0983= 0.6419 → 64.2 %

6.4 % + 29.4 % +64.2 % = 100% = 1

This results will be shown to the decision maker, which will validate his choice after that to be stored as models in the Data Base. this validated decisions will be considered as models, so the next time the system will consider the models and it will govern differently for example if this time results give that the best alternative is number one and the user don't choose it, in the next time if same alternatives (or similar) have been entered, alternative one will not have the highest percentage. So the system will give better results (According to the user's desire sure) than the last time. Same procedure in each operation will ameliorate the system as user want and to be commensurate with his desire, and this is one of the characteristics of intelligent system which is perceive, learning autonomously.

4.3 Clustering

In the previous section we describe how we use naïve Bayes classifier. As shown above it is used for classification, we have two previously defined classes *accepted* and *denied* (accepté / refusé) and we want to classify the new instances. The problem if you don't know what the classes are? And how many classes you have? When you don't have a defined classes there is a technique called clustering (also called unsupervised learning) apply when there is no class to be predicted but rather when the instances are to be divided into natural Groups.

A Cluster is a group of instances having the same characteristics, also can be defined as a subset of data which are similar. Clustering is the process of dividing a dataset into groups such that the members of each group are as similar (close) as possible to one another, and different groups are as dissimilar (far) as possible from one another. Clustering can uncover previously undetected relationships in a dataset.

How to determine the similarity between two objects represent an important issue in clustering, so objects with high similarity within clusters and low similarity between clusters can form the clusters. Commonly, to measure similarity or dissimilarity between objects, a distance measure such as Euclidean, Manhattan, Minkowski (for numeric data) and Jaccard coefficient (binary data) is used. A distance function returns a stronger value for pairs of objects that are less similar.

many functions (distance) used for the similarity measure:

Euclidean

$$\sqrt{\sum_{i=1}^{k} (\mathbf{x}_i - \mathbf{y}_i)^2}$$

Manhatan

$$\sum_{i=1}^{k} |\mathbf{x}_i - \mathbf{y}_i|$$

Minkovski

$$\left(\sum_{i=1}^{k} (|\mathbf{x}_i - \mathbf{y}_i|)^q\right)^{1/q}$$

There is two types of clustering: hierarchical and partitive. In the hierarchical clustering there is agglomerative method and divisive method. For partitive clustering there is K-Means algorithm and self-organizing map (fig 3.9)

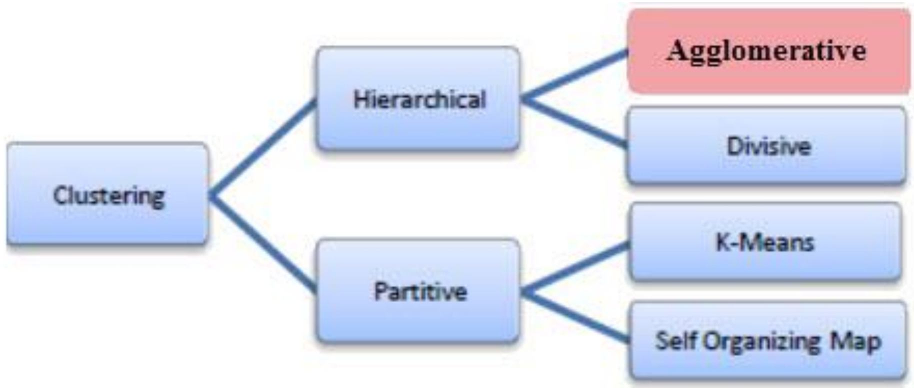


Fig 3.5 clustering types in data mining [Sayad, 2016]

Divisive method (top-down method)

The principle of this method, rests primarily to assign all observations to a single cluster, after this, divide the cluster into two less similar clusters. The goal is to reach a cluster for each observation, so the procedure of dividing each cluster into two cluster, will be recursively until the goal is reached.

Agglomerative method (bottom-up method)

Unlike the divisive method, agglomerative method is based on four main stages. The first is to assign each observation to its own cluster. The second step is to calculate the similarity (eg, distance) between each group. The third step is to join the two most similar groups. The fourth and the final stage is a repeat of steps 2 and 3 until there is only one cluster left. **[JOHNSON, 67]** propose an algorithm for this method.

As we mentioned previously, we propose to manage the database of citizens. In the data base we have the public participation as shown in the architecture (fig 3.1), which means that we have the vote of each citizen (satisfied/dissatisfied/indifferent) for the projects vote which participate. From the beginning we said that the participation of public should be with constraints, because there is always people that could undermine the operation. Our idea here is to categorize the citizens depending on their participation, active citizens (their participation varies: satisfied, dissatisfied, indifferent…), passive citizens (always dissatisfied) …and after that we eliminate the passive citizens. To make sure that we will hurt anyone with the elimination, we take just the participants who attended over the threshold (ten or more votes).

We will use the agglomerative method to get the different categories (clusters). As we have binary data which varies between satisfied and dissatisfied (Yes/No) (we have also indifferent but it's considered as no). And the attributes (projects) that we have are asymmetric. The first step is building the contingency table (also called dissimilarity table), it's based on four dependent quantities:

a = proportion of 1s that the variables share in the same positions

b = proportion of 1s in the first variable and 0s in second variable in the same positions

c = proportion of 0s in the first variable and 1s in second variable in the same positions

d = proportion of 0s that both variables share in the same positions.

[WARRENS &AL, 08]

	Variable two		
Variable one	Value 1	Value 0	Total
Value 1	a	b	p_1
Value 0	c	d	q_1
Total	p_2	q_2 **Sum**	p_1 p_2 q_1 q_2

Fig 3.6 contingency table [WARRENS & AL, 08]

Jaccard coefficient (for the binary variable is asymmetric)

$$\mathrm{D}\boldsymbol{Jac} = \frac{b + \mathrm{c}}{\mathrm{a} + \mathrm{b} + \mathrm{c}}$$

Considering this citizens table:

Project / Citizen	Project1	Project2	Project3	Project4	Project5
Citizen_1	Y	Y	Y	Y	N
Citizen_2	Y	Y	Y	Y	Y
Citizen_3	N	N	Y	Y	Y
Citizen_4	N	Y	Y	N	N
Citizen_5	N	N	N	N	N
Citizen_6	N	N	N	N	N

Tab 3.6 example of citizens table

Let's take a deeper look at how we apply Johnson's algorithm in our project.

NB: this example is application of Johnson's algorithm in case of single-linkage clustering.

Begin with calculating the distance between citizens using *jaccard coefficient.*

		Citizen_2		
		1	0	Sum
Citizen_1	1	4	0	4
	0	1	0	1
	Sum	5	0	5

Tab 3.7 Contingency table (Citizen_1 and Citizen_2)

$$D\,(\text{Citizen}_1, \text{Citizen}_2) = \frac{b + c}{a + b + c} = \frac{0 + 1}{4 + 0 + 1} = \frac{1}{5} = 0.2$$

The distance between Citezen_1 and Citizen_2 is: 0.2

		Citizen_2		
		1	0	Sum
Citizen_1	1	2	2	4
	0	1	0	1
	Sum	5	0	5

Tab 3.8 Contingency table (Citizen_1 and Citizen_3)

$$D\ (\text{Citizen}_1, \text{Citizen}_3) = \frac{2+1}{2+2+1} = \frac{3}{5} = 0.6$$

The distance between Citezen_1 and Citizen_3 is: 0.6

$$D\ (\text{Citizen}_1, \text{Citizen}_4) = \frac{2+0}{2+2+0} = \frac{2}{4} = 0.5$$

$$D\ (\text{Citizen}_3, \text{Citizen}_4) = \frac{2+1}{1+2+1} = \frac{3}{4} = 0.75$$

$$D\ (\text{Citizen}_5, \text{Citizen}_6) = 0$$

Citizen	Citizen_1	Citizen_2	Citizen_3	Citizen_4	Citizen_5	Citizen_6
Citizen_1	0	0.2	0.6	0.5	1	1
Citizen_2		0	0.4	0.6	1	1
Citizen_3			0	0.75	1	1
Citizen_4				0	1	1
Citizen_5					0	0
Citizen_6						0

Tab 3.9 Distance table

The shortest distance between pair of citizens is Citizen_5 and Citizen_6, at distance 0 (Similar).so we grouped them into a single cluster "Citizen_5 / Citizen_6" and we take the shortest distance for each Citizen (Fig 3.7).

Step 1

Citizen	Citizen_1	Citizen_2	Citizen_3	Citizen_4	Citizen_5 / Citizen_6
Citizen_1	0	0.2	0.6	0.5	1
Citizen_2		0	0.4	0.6	1
Citizen_3			0	0.75	1
Citizen_4				0	1
Citizen_5 / Citizen_6					0

Tab 3.10 group Citizen_5 and Citizen_6

Step 2

Citizen	Citizen_1 / Citizen_2	Citizen_3	Citizen_4	Citizen_5 / Citizen_6
Citizen_1 / Citizen_2	0	0.4	0.5	1
Citizen_3		0	0.75	1
Citizen_4			0	1
Citizen_5 / Citizen_6				0

Tab 3.11 group Citizen_1 and Citizen_2

Step 3

Citizen	**Citizen_1 / Citizen_2 / Citizen_3**	**Citizen_4**	**Citizen_5 / Citizen_6**
Citizen_1 / Citizen_2/ Citizen_3	0	0.5	1
Citizen_4		0	1
Citizen_5 / Citizen_6			0

Tab 3.12 group Citizen_3 and Cluster1

Step 4

Citizen	**Citizen_1 / Citizen_2 / Citizen_3 / Citizen_4**	**Citizen_5 / Citizen_6**
Citizen_1 / Citizen_2/ Citizen_3/ Citizen_4	0	1
Citizen_5 / Citizen_6		0

Tab 3.13 group Citizen_4 and Cluster1

In the end we get to clusters: (Citizen_1, Citizen_2, Citizen_3, Citizen_4) active citizens (Citizen_5, Citizen_6) passive citizens.

So this is how we divide the citizens to get the passive ones and eliminate them or just ask them, if they are sure of their votes?, because it's abnormal to be dissatisfied for all the projects especially the number is huge.

NB

What we have seen in this chapter (naïve Bayes demarche and clustering) are just examples that we realized manually (not the system) to understand the contribution. The datasets much more attributes and much more instances. The next chapter contains more details.

5 Conclusion

We have discussed in this chapter our contribution which was composed of an architecture and there principal components and processes, the demarche of such process. In the next chapter we will introduce the tools that we used to develop the system and some pictures of it.

Chapter 4

Implementation and experimentation of the proposed architecture

1 Introduction

In this chapter we presents the tools and the frameworks that we have familiarized with and that we have used to realize, develop the system (prototype) such as the big data tools, the used data, the operating system...and the obtained results after experimentation and see if it's conform with the theoretical results seen in the previous chapter.

2 Development tools

2.1 Cloudera

Cloudera was the first, and is currently, the leading provider and supporter of Apache Hadoop for the enterprise, it was founded in 2008. Cloudera offers software for business critical data challenges including storage, access, management, analysis, security, and search.

The first unified Platform for Big Data was offered by Cloudera which is The Enterprise Data Hub. It was really a revolution in enterprise data management. **[Cloudera, 2016]**

Cloudera also offers students, developers and researchers that can't work in distributed mode the opportunity to develop and work in standalone mode (master and slaves in the same machine) and offers them Cloudera virtual machine CDH[3].

CDH[3] : Cloudera Distribution Including Apache Hadoop

2.2 Cloudera Virtual Machine

Also called Cloudera distribution of Apache Hadoop (CDH), The Cloudera distribution of Apache Hadoop and other related open-source projects, including Cloudera Impala and Cloudera Search. CDH also provides security and integration with numerous hardware and software solutions. Cloudera has released many version (the last is CDH 5.7, April 2016) and we have used in the development CDH 5.3.0 which work with CentOS Operating system, which means that, as in all Linux Operating systems (Ubuntu, Red Hat, SUSE, Debian, Mandriva …), all operations realized in command line.

Fig 4.1 Cloudera's desktop

2.2.1 Hadoop eco-system

The Apache™ Hadoop® project develops open-source software for reliable, scalable, distributed computing.

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets (Big Data) across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures. **[HADOOP, 2016]**

The project includes these modules:

Hadoop Distributed File System (HDFS™): A distributed file system that provides high-throughput access to application data.

Hadoop YARN: A framework for job scheduling and cluster resource management.

Hadoop MapReduce: Effective implementation of the MapReduce algorithm and A YARN-based system for parallel processing of large data sets.

Hadoop Common: The common utilities that support the other Hadoop modules.

Hadoop provides a series of tools ready to use **[HADOOP, 2016]**:

HBase™: A scalable, distributed and non-relational database that supports structured data storage for large tables.

ZooKeeper™: A high-performance coordination service for distributed applications.

Hive™: A data warehouse infrastructure that provides data summarization and ad hoc querying with a query language similar to SQL.

Mahout™: A Scalable machine learning and data mining library.

Pig™, Spark™, Ambari, **Avro™**, **Cassandra™**, **Chukwa™**, **Tez™**, [HADOOP, 2016]

Fig 4.2 Hadoop version from terminal

2.2.2 HBase

Our Data Base that appear in the architecture and used to store models is HBase. HBase provides random, real-time read/write access to Big Data. This project's goal is the hosting of very large tables -- billions of rows X millions of columns -- atop clusters of commodity hardware. Apache HBase is an open-source, distributed, versioned, non-relational database modeled after Google's Bigtable: A Distributed Storage System for Structured Data by **[CHANG ET AL, 06]**. Just as Bigtable leverages the distributed data storage provided by the Google File System, Apache HBase provides Bigtable-like capabilities on top of Hadoop and HDFS.

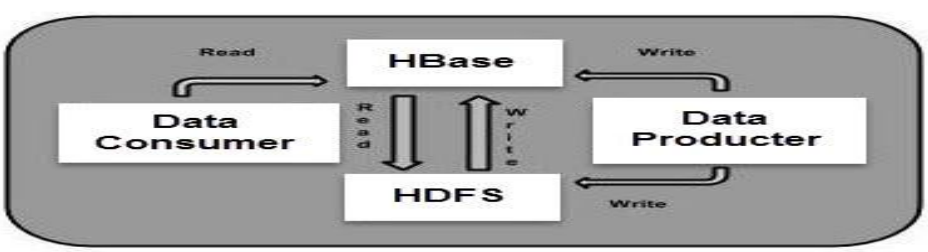


Fig 4.3 HBase read/write data from/in HDFS

There are three major components to HBase: the client library, one master server, and many region servers. The region servers can be added or removed while the system is up and running to accommodate changing workloads. The master is responsible for assigning regions to region servers and uses *Apache* ZooKeeper, a reliable, highly available, persistent and distributed coordination service, to facilitate that task. HBase uses ZooKeeper also to ensure that there is only one master running, to store the bootstrap location for region discovery, as a registry for region servers, as well as for other purposes. ZooKeeper is a critical component, and without it HBase is not operational. [GEORGE, 2011]

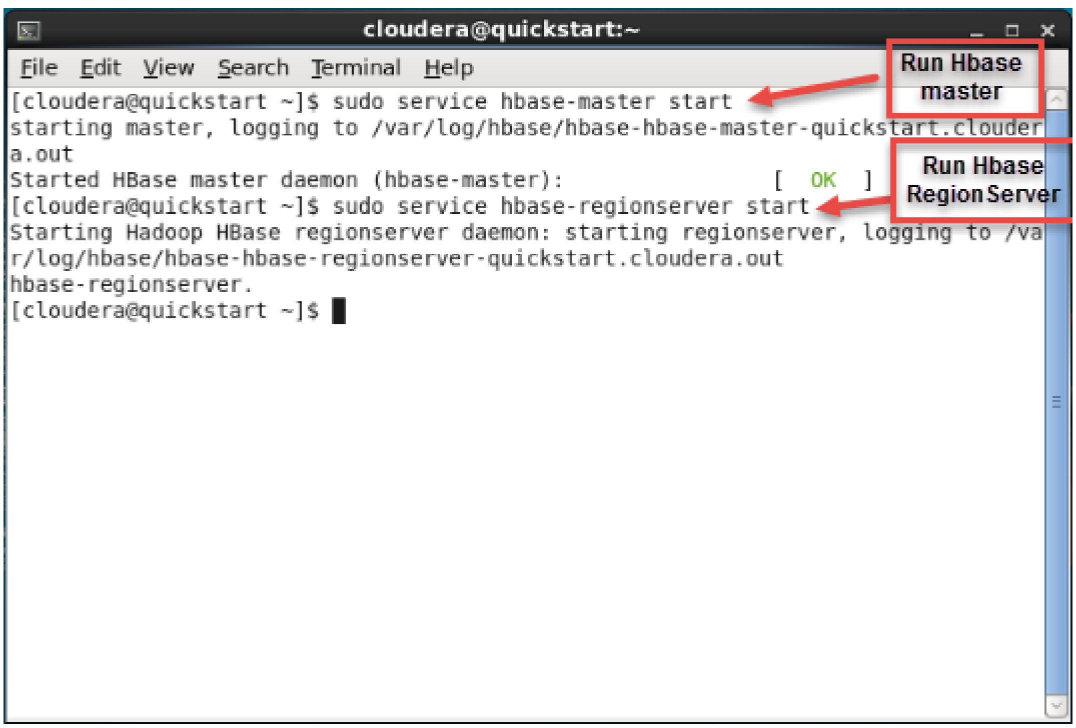
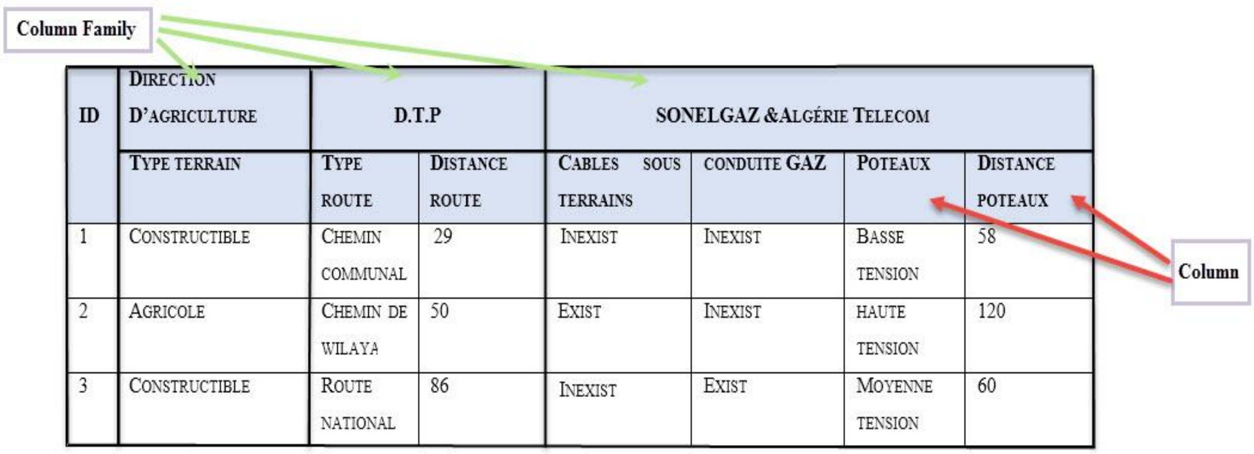


Fig 4.4 run HBase master and RegionServer

In HBase:

- Table is a collection of rows.
- Row is a collection of column families.
- Column family is a collection of columns (one or many).
- Column is a collection of value.



ID	Direction d'agriculture	D.T.P		SONELGAZ &Algérie Telecom			
	Type terrain	Type route	Distance route	Cables sous terrains	conduite GAZ	Poteaux	Distance poteaux
1	Constructible	Chemin communal	29	Inexist	Inexist	Basse tension	58
2	Agricole	Chemin de wilaya	50	Exist	Inexist	haute tension	120
3	Constructible	Route national	86	Inexist	Exist	Moyenne tension	60

Fig 4.5 Example of HBase table

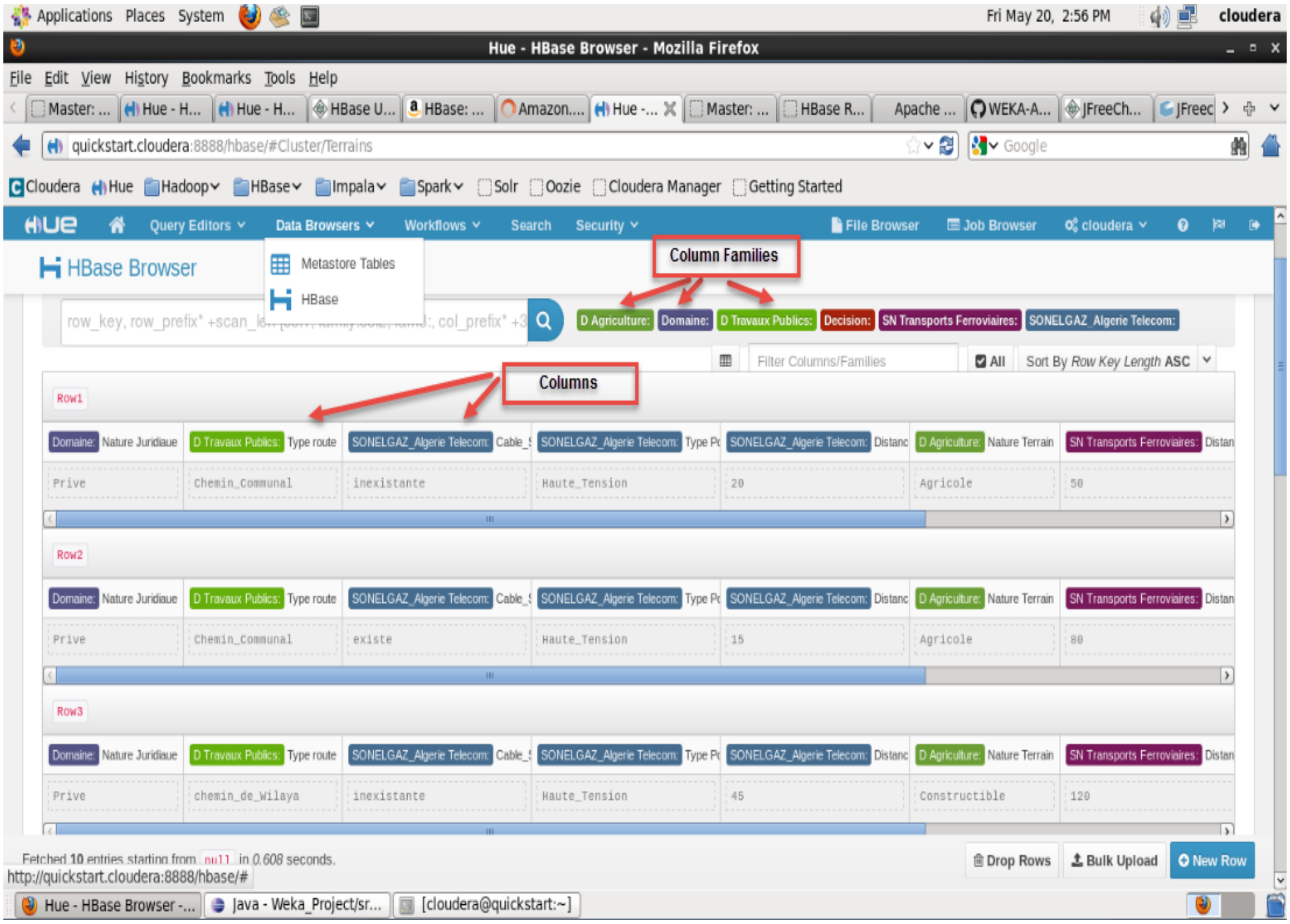


Fig 4.6 Table of fields in HBase

We use Hue[4] to visualize the table.

Advantages of HBase:

- Sits on top of the HDFS provides random real-time read/write access to data in the Hadoop File System. (Fig 4.3).

Hue[4]: Hue is a Web interface for analyzing and visualizing data with Apache Hadoop

- HBase have a better degree of fault tolerance for large-scale data management in some sense. Because DFS provides reliable storage, and current database systems cannot make use of DFS directly. **[ZHANG ET AL, 2010]**
- A scalable storage solution to accommodate a virtually endless amount of data. [GEORGE, 2011]
- HBase scales to billions of rows and millions of columns. [GEORGE, 2011]
- It's possible also to import other data bases (RDB[5] for example) in HBase, that's will suitable for us.

For those reasons we believe that HBase is the best solution to store our data.

NB:

It should be noted that we don't have really big data in this implementation (till writing this words), but we use a big data tool (HBase). Although the stored data not voluminous (what we managed to collect), it subsequently becomes. In every planning we have quantity of data that will be stored (the alternatives).how many urban project launched per week? Per month? Per year?

In addition to the huge volume of data (see Chapter 3 section 3.3) stored by the administrations (members of urban planning stuff) in different domains, you can imagine now the quantity of data that will be stored in our Data Base. It's never ever possible to control this mass any more with traditional DBMS[6].

RDB[5]: Relational Data Base
DBMS[6]: Data Base Management System

2.2.3 Eclipse

Eclipse Foundation produce and provide tools for the realization of software, including programming activities including integrated development environment(IDE) and frameworks, and also other activities covering modeling, design, testing, configuration management, reporting. Cloudera contains the IDE eclipse Juno and we used it in the development. Its IDE part of the project aims to support any programming language C++, **Java…** this last is a pure object-oriented programming language.

2.2.4 VMware Workstation 12 Pro

VMware Workstation virtual machine monitor also called hypervisor that offers the possibility to create and run virtual machines. VMware delivers primary edge features and performance that technical professionals rely on every day when working with virtual machines. With support for the latest version of Windows and Linux, the latest processors and hardware, it's the perfect tool to run Virtual Machines. (Last version 12.1.1 / 21 April 2016) **[VMWARE, 2016]**

2.2.5 Weka

Other tool that we have used is weka. It is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code and it is possible to apply it to big data. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. Weka is open source software issued under the General Public License (GNU). **[HALL ET AL, 09]**

3 Study case implementation

In this section we show the important interfaces of the implemented use case. We can called it prototype, because it's not the last version, it need ameliorations but it's functional. Before that we present's our system characteristics:

Dell - Inspiron 15 Laptop

- Intel Core i5, 2 virtual CPU cores given to cloudera.

- 8GB Memory, we configure Cloudera with 6GB.

- 1TB Hard Drive

3.1 Make prediction (Naïve Bayes)

The first functionality of the system is to make prediction of decisions.

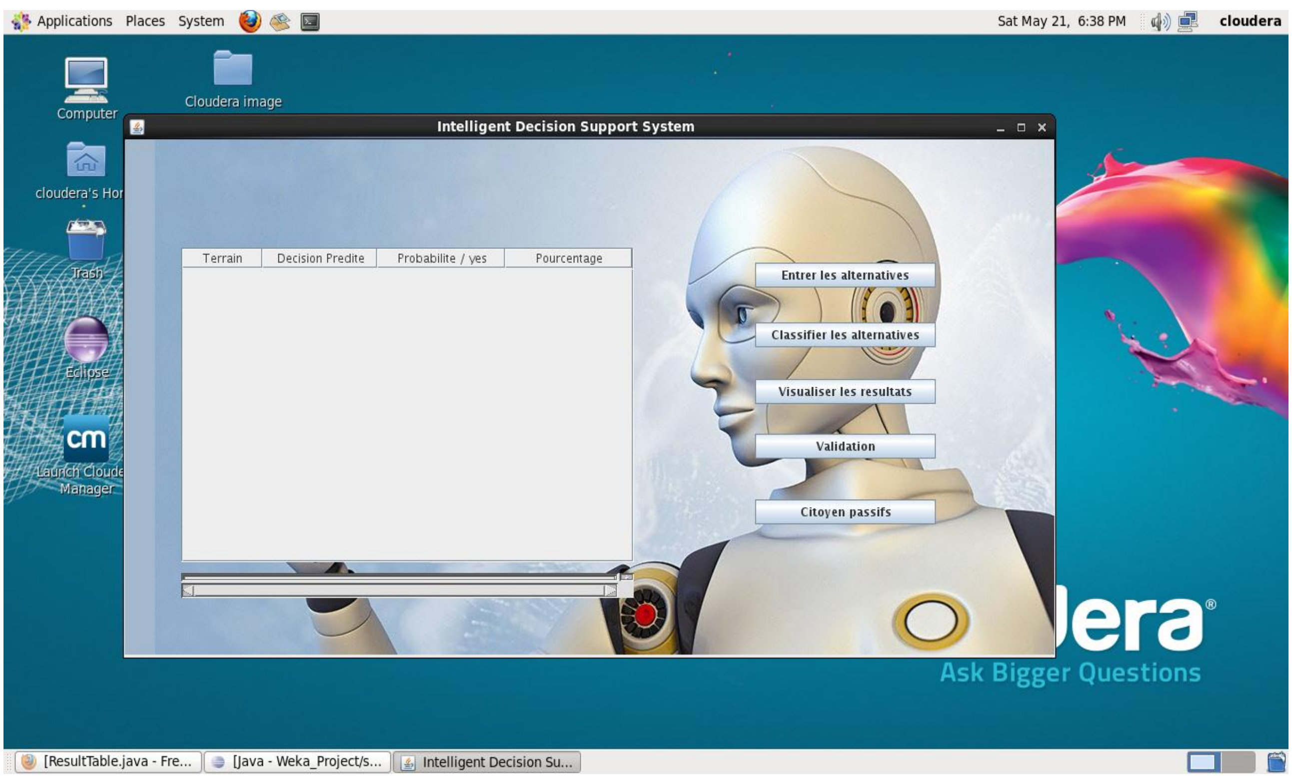


Fig 4.7 Principal window

Entering the alternatives which will be classified 3 fields with different characteristics.

Fig 4.8 the new operation window

Classify alternatives (naïve Bayes) and get the results:

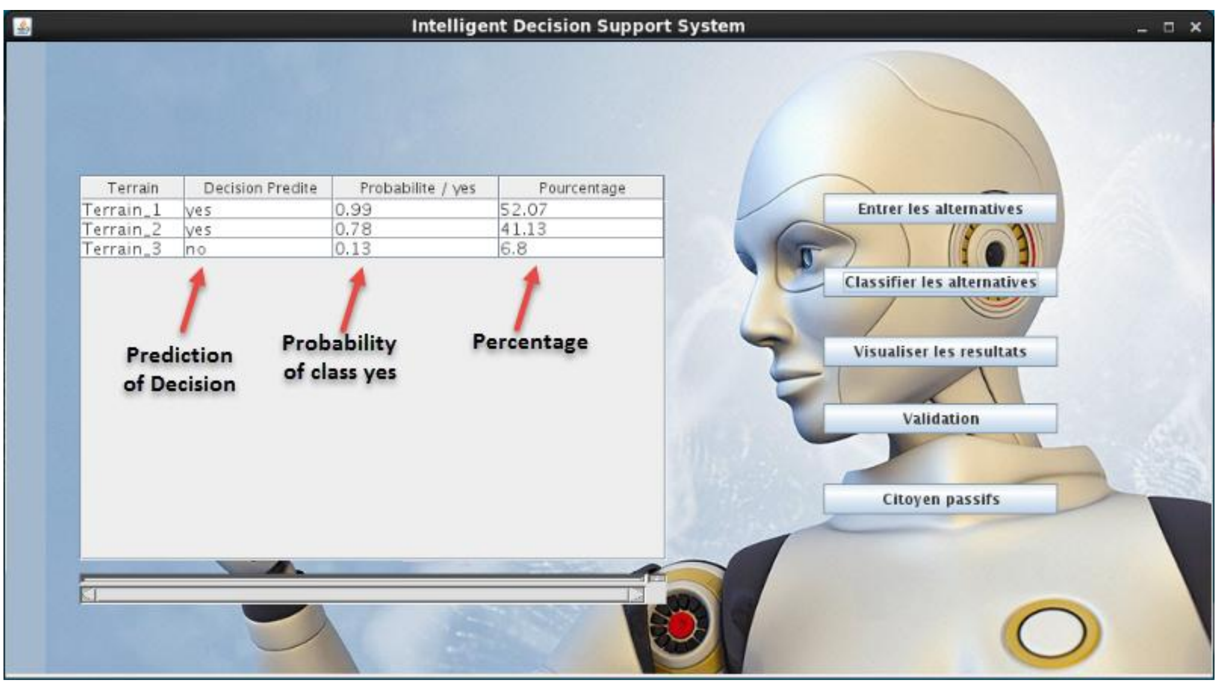


Fig 4.9 Results of classification

Visualization of the results in charts:

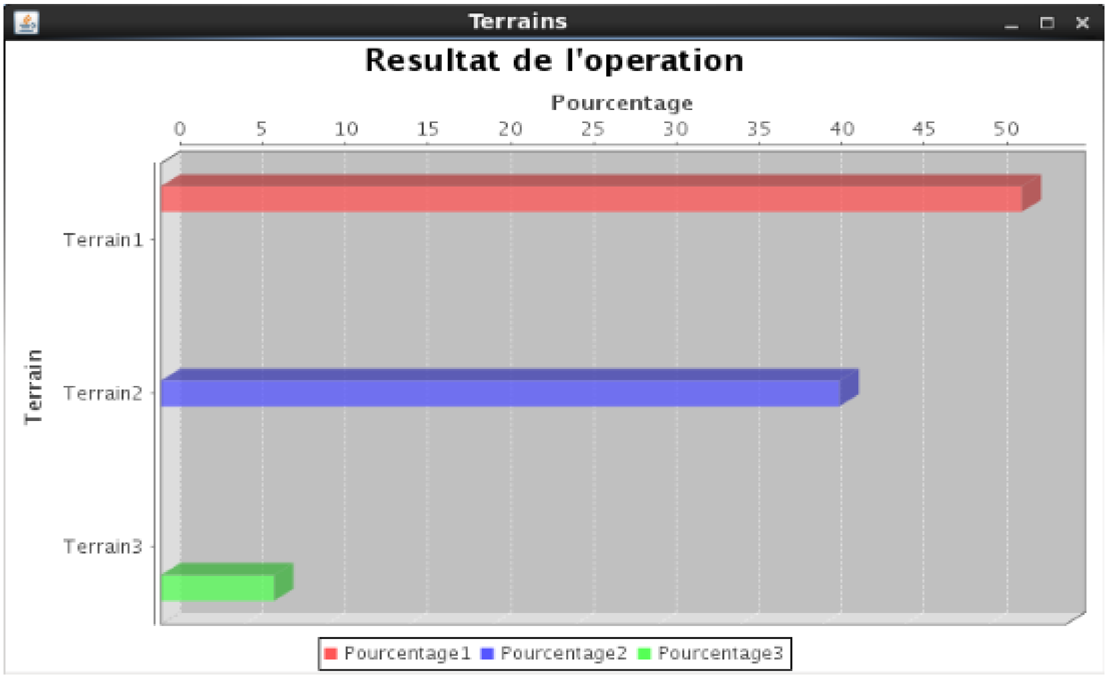


Fig 4.10 Results in BarChart

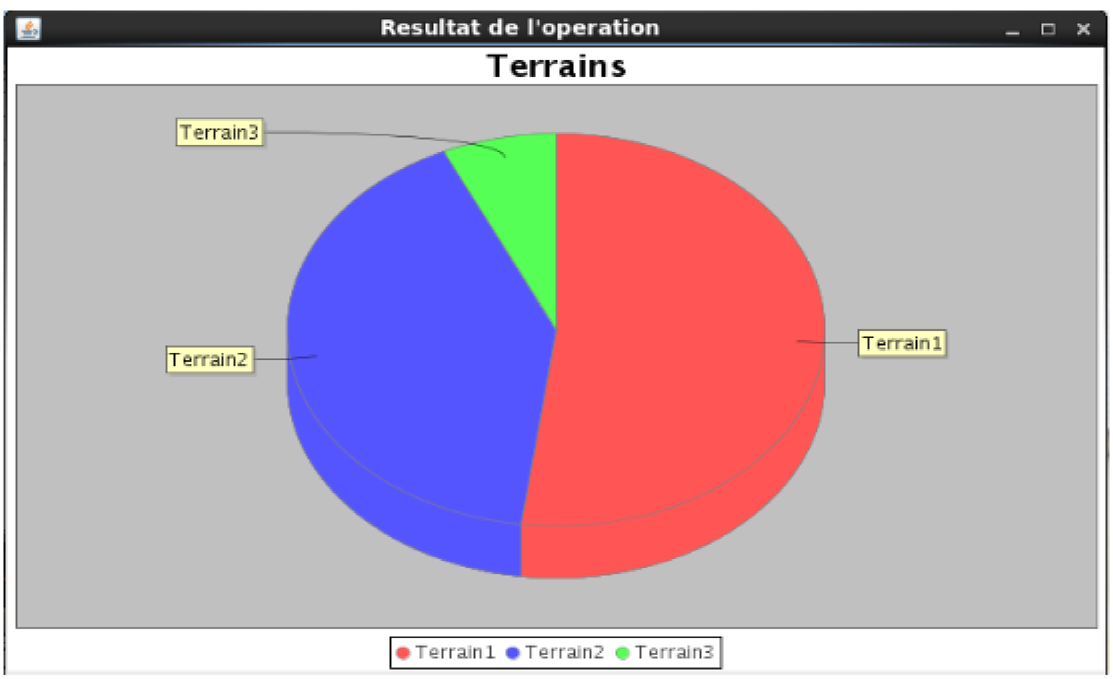


FIG 4.11 Results in PieChart

Best alternative according to the system is the first. The system will ask to validate the choice, or the user do it voluntary. So what if we don't choose the first alternative (which is the best) and validate other alternative, how will the system react then? Let's see...

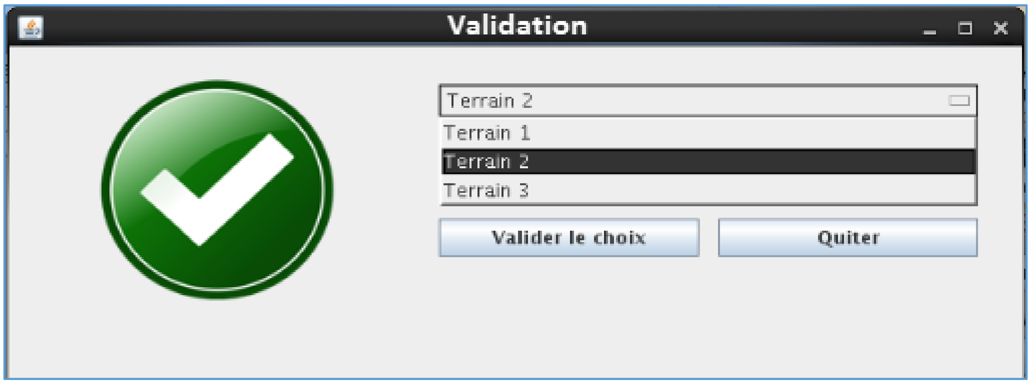
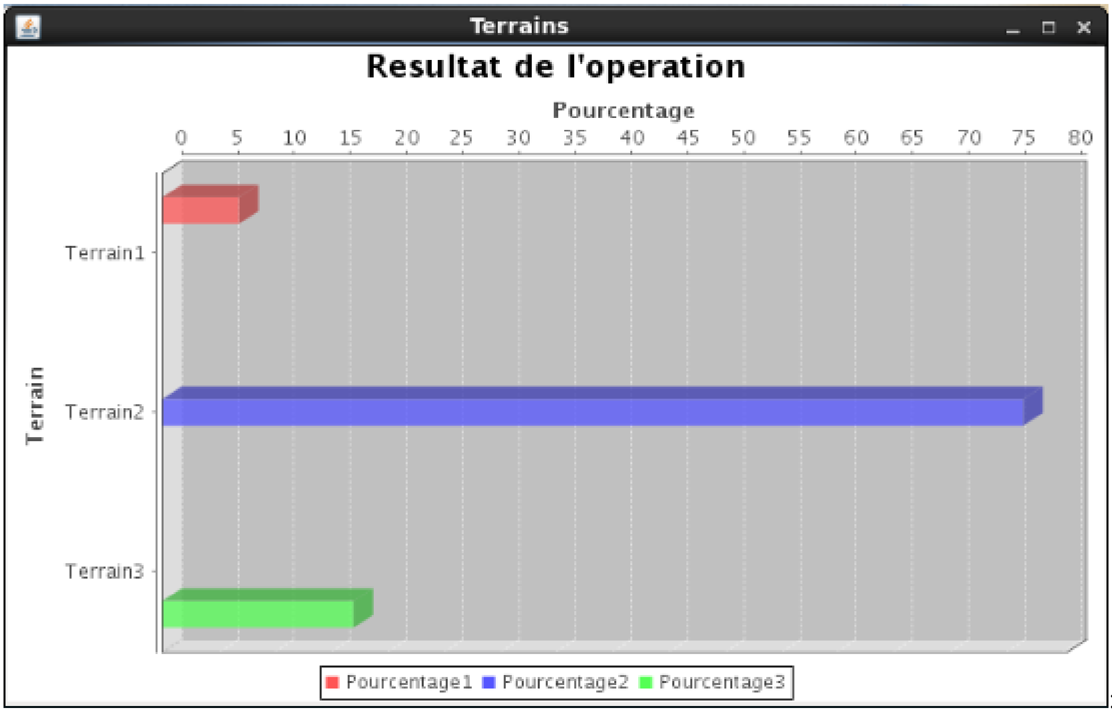


FIG 4.12 validation of choice

We have validated second choice which have not the highest percentage. All the alternatives (that already classified) are in the Model Base. Let's reenter the same alternatives (same characteristics) and get the result again.

Fig 4.13 reenter the same alternatives

Classify the alternatives and get the results



:

Fig 4.14 Results after validation in BarChart

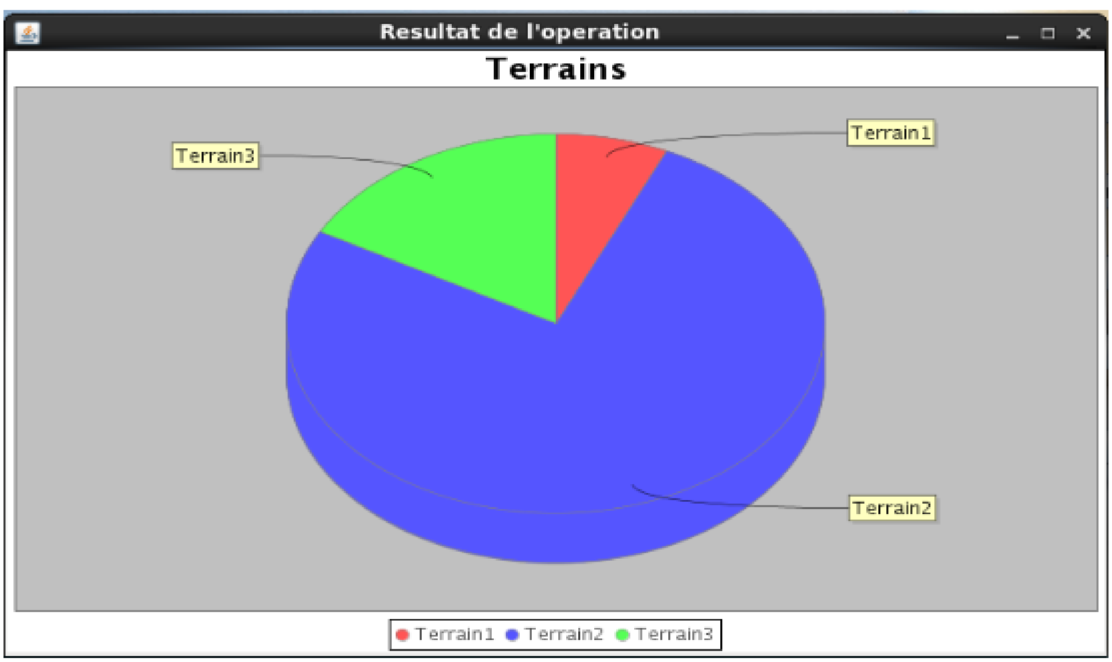


Fig 4.15 Results after validation in PieChart

The system change the result according to our tendencies and desires, so the theoretical results that we got in the previous chapter are correct.

Although not the same examples (we don't have same number of attributes) but behavior of the system changed as we predict it.

3.2 Passive citizens (Clustering)

As we don't have really a citizens participations, we have made a simple dataset to be learned.

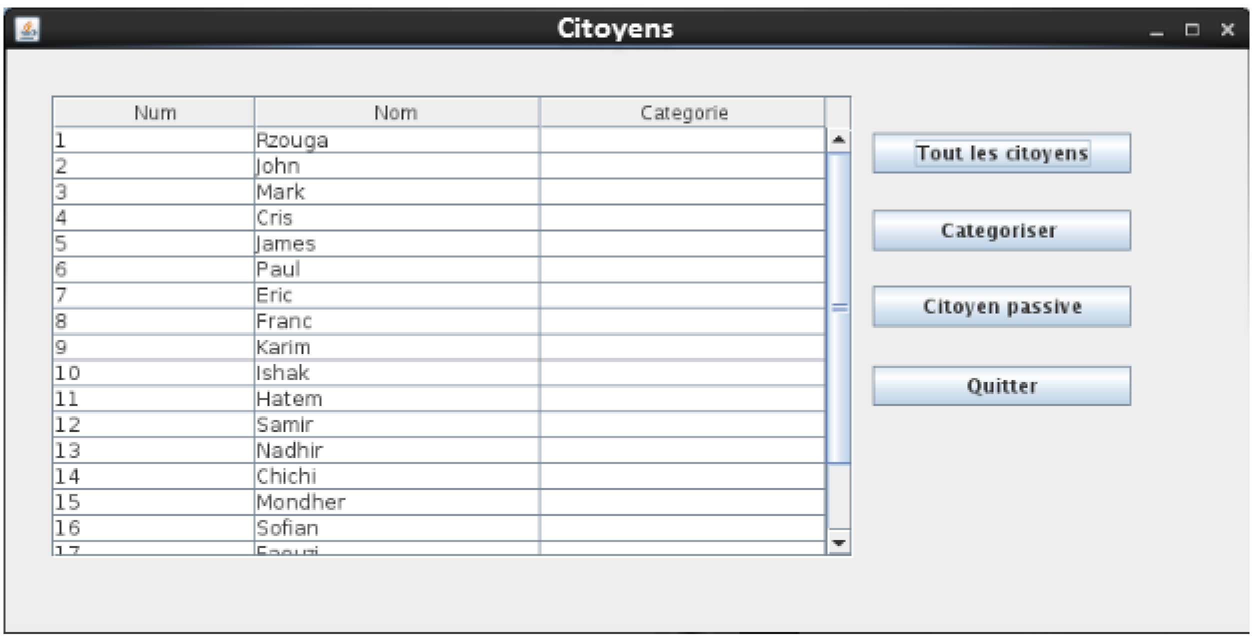


Fig 4.16 Citizens window (All the citizens)

Clustering all citizens according to their votes:

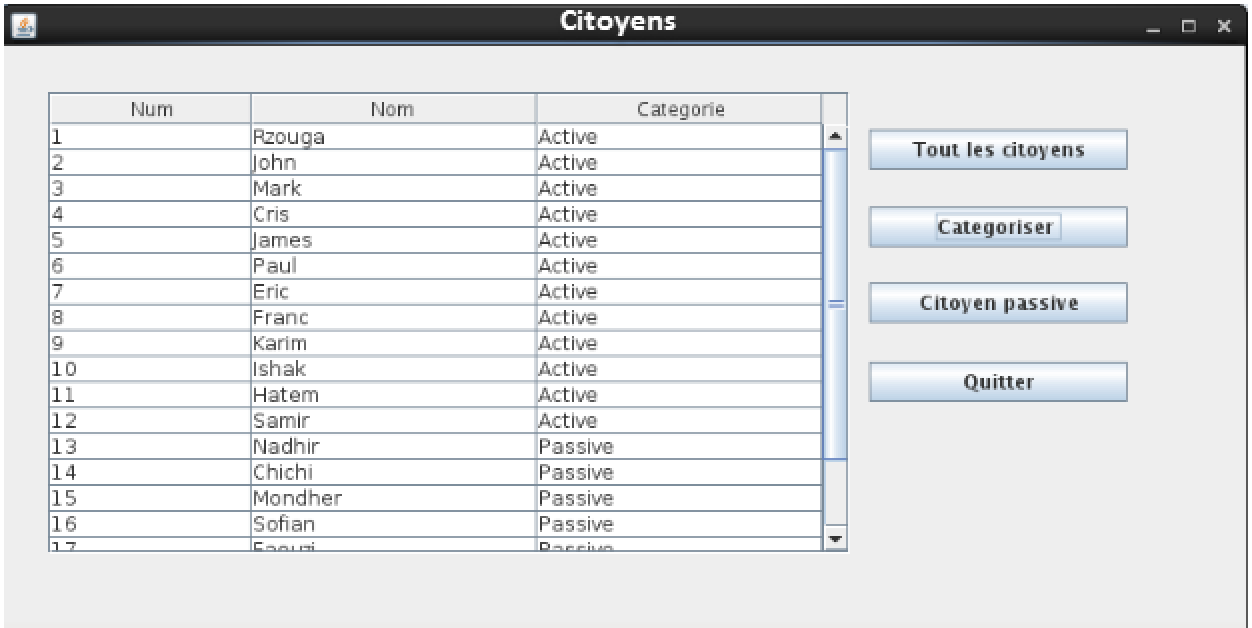


Fig 4.17 Citizens categorization

Get only the passive citizens:

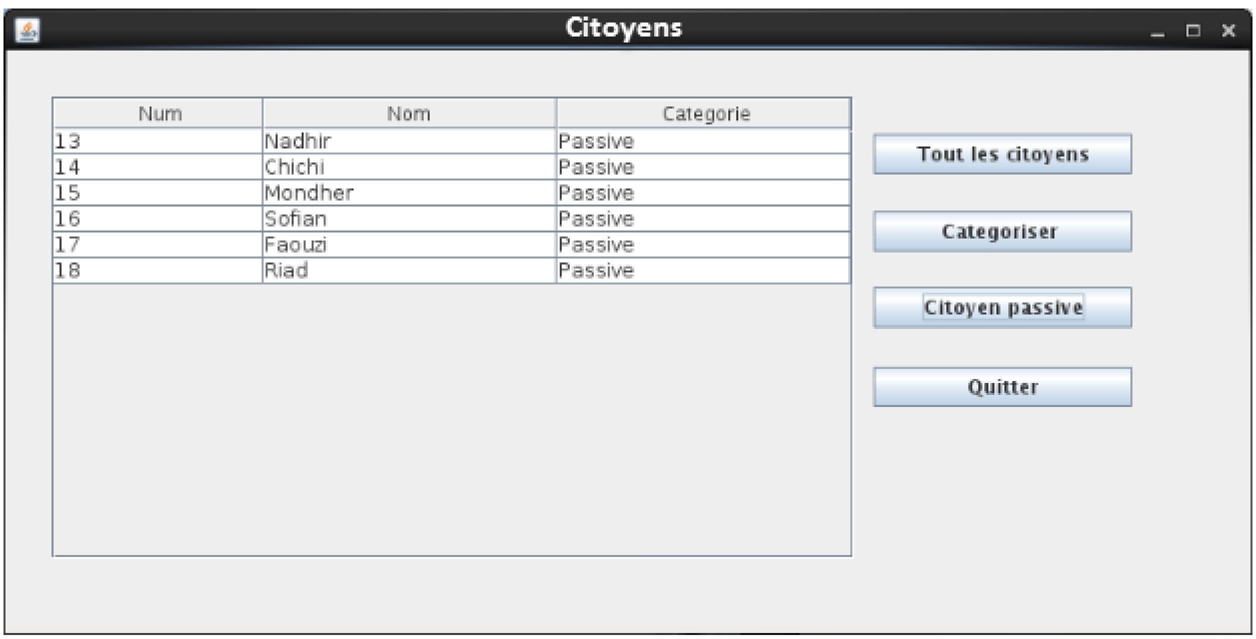


Fig 4.18 Passive Citizens

4 Conclusion

We have presented in this chapter our proposed prototype and the used data relative to the future system we have applied a use case consisting of real collected data. Finaly the obtained results are considered very encouraging and very interesting and showed accordingly efficiency and robustness of the proposed architecture.

General conclusion and perspectives

The clear sign of the complexity of the decision in urban planning. The existence of multiple easements (criteria) used that has several actors (administrations), and each project realization, there are different consequences which can be excellent, good, poor or bad.

In the problematic we handle, which is the choice of land that must introduce the interdependencies between actors and policy makers, and citizens who represent a fundamental element in our proposed process (Reservedly).

Urban planning is usually multi-criteria. where several conflicting criteria are often taken into account in evaluating different Planning process which present a problem for the meter agree to a solution that satisfy all participants, therefore traditional system are no longer able to assist policymakers. Consequently we proposed architecture of intelligent system that have the ability to solve the cited problems and offers better performance.

Thus, we have initiated within the domain of intelligent spatial decision support a new architecture combining Data mining techniques:

First: Naïve bayes, to make prediction of decision whether accepted or denied, in addition classification of the alternatives. Also using outcomes to ameliorate results.

Second: Clustering, to categorize the participants of citizens in process to get those can mess with the planning operation.

Third: Big data tools, to manage the huge amount of data which can't be managed with relational data base management systems (RDBMS).

Perspectives

- Adapting our architecture with cloud computing to beneficiate from this powerful tool.
- Ameliorate the technique that we use to detect passive citizens.
- Use other artificial technique methods which may be more preferment.
- Complete the other part (participation of citizens) of architecture that may use other approaches, other tools and other techniques especially mobile applications which can be our solutions to manage public participations

References:

[AKINOLA ET AL, 2015] AKINOLA, S. Olalekan et OYABUGBE, O. Jephthar. Accuracies and Training Times of Data Mining Classification Algorithms: An Empirical Comparative Study. Journal of Software Engineering and Applications, 2015, vol. 8, no 9, p. 470.

[ARMSTRONG ET AL, 1986] Armstrong, M. P., Densham, P. J. and Rushton, G., “Architecture for a microcomputer based spatial decision support system,” Second International Symposium on Spatial Data Handling, 120, 131 International Geographics Union, 1986.

[ASADI, 2015] ASADI, Reza et GHATEE, Mehdi. A Rule-Based Decision Support System in Intelligent Hazmat Transportation System. Intelligent Transportation Systems, IEEE Transactions on, 2015, vol. 16, no 5, p. 2756-2764.

[BRANS & VINCKEJ, 85]_P. BRANS AND PH. VINCKEA PREFERENCE RANKING ORGANISATION METHOD (The PROMETHEE Method for Multiple Criteria Decision-Making) MANAGEMENT SCIENCE Vol. 31, No. 6, June 1985Printed in U.S.A.

[BRANS & MARESCHAL, 05] BRANS, Jean-Pierre et MARESCHAL, Bertrand. PROMETHEE methods. In :Multiple criteria decision analysis: state of the art surveys. Springer New York, 2005. p. 163-186.

[BERKHIN, 06]_BERKHIN, Pavel. A survey of clustering data mining techniques. In: Grouping multidimensional data. Springer Berlin Heidelberg, 2006. p. 25-71.

[BUTLIN, 89] Butlin, John (1989-04-01). "Our common future. By World commission on environment and development. (London, Oxford University Press, 1987, pp.383)". Journal of International Development 1 (2): 284–287.

[BOUYSSOU ET AL, 02]Denis Bouyssou, Eric Jacquet-Lagrèze, Patrice Perny, Roman Slowiński, Daniel Vanderpooten and P. Vincke. Aiding Decisions with Multiple Criteria: Essays in Honor of Bernard Roy. Springer Science+Business Media New York 2002

[BELTON & STEWART, 02] Valerie Belton, Theodor J Stewart Multiple Criteria Decision Analysis An Integrated Approach Kluwer Academic Publishers in 2002 Softcover reprint of the hardcover 1st edition 2002

[BURSTEIN & HOLSAPPLE, 08] Burstein, F., & Holsapple, C. W. New horizons at the core of information systems: decision support system advances. *Information Systems and E-Business Management*, *6*(2), 109-115.

[BRANS ET AL, 75] BRANSJ, . P., B. RoY AND Ph. VINCKE, "Aide a la decision multicritere," Rev. Belge Statisti., d'Informatique et de Recherche Operationnelle, 15, 4 (1975).

[CLOUDERA, 2016]______http://www.cloudera.com/documentation/enterprise/5-3-x/topics/introduction.html Page generated April 22, 2016.

[CASTELLS, 77] Castells M (1977) The Urban Question. English edition. Edward Arnold.

[CALZADA ET AL, 2015] CALZADA, Alberto, LIU, Jun, WANG, Hui, et al. Application of a Spatial Intelligent Decision System on Self-Rated Health Status Estimation. Journal of medical systems, 2015, vol. 39, no 11, p. 1-18.

[CORTÉS ET AL., 2000] CORTÉS, Ulises, SÀNCHEZ-MARRÈ, Miquel, CECCARONI, Luigi, et al. Artificial intelligence and environmental decision support systems. Applied Intelligence, 2000, vol. 13, no 1, p. 77-91.

[COVINA ET AL, 01] Covina, J.G., Slevin, D.P., Heeley, M.B. Strategic decision making in an intuitive vs. technocratic mode: structural and environmental considerations. Journal of Business Research (52) 2001, pp 51-67.

[CROSSLAND ET AL 1995] Crossland, M. D., Wynne, B. E. and Perkins, W. C., “Spatial Decision Support Systems: An overview of technology and a test of efficacy,” Decision Support Systems, 14, 3, 1995, 219-235.

[CHAKHAR ET MOUSSEAU, 07] Salem Chakhar, Vincent Mousseau MULTICRITERIA

SPATIAL DECISION SUPPORT SYSTEMS 2007

[CHAKHAR & MARTEL, 03] S. Chakhar and J.-M. Martel, Enhancing geographical information systems capabilities with multi-criteria evaluation functions. Journal of Geographic Information and Decision Analysis, Vol. 7, No. 2, pp. 47-71, 2003.

[CHAKHAR & MOUSSEAU, 09] CHAKHAR, Salem et MOUSSEAU, Vincent. Generation of spatial decision alternatives based on a planar subdivision of the study area. In: Advanced Internet Based Systems and Applications. Springer Berlin Heidelberg, 2009. p. 137-148.

[DHAR & STEIN 97] Dhar, V., Stein, R.: Intelligent Decision Support Methods: The Science of Knowledge Work. Prentice-Hall Inc, Upper Saddle River (1997)

[Doubriere, 79] Doubriere JC (1979) Cours d'urbanisme appliqué, Paris: Eyrolles, 260 pp.

[DANIELA ET AL, 09] Daniela, X., Christopher, J.H. and Roger, G.S. (2009) Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. *IJSCI: International Journal of Computer Science Issues*, **4**.

[FALUDI, 73] Faludi A (1973) *Planning Theory.* Oxford: Pergamon Press. Farley J and Hecht L (1999) *OGC Discussion Paper-Disaster Management Scenarios,* Open GIS Consortium, January, 1999.

[FIGUEIRA ET AL, 05] J. Figueira, S. Greco and M. Ehrgott, Multiple Criteria Decision Analysis: State of the Art Surveys. Springer-Verlag, New York, 2005.

[FAYYAD ET AL, 96] Fayyad, U. M.; Piatetsky-Shapiro, G.; and Smyth, P. 1996. From Data Mining to Knowledge Discovery: An Overview. In *Advances in Knowledge Discovery and Data Mining,* eds. U. Fayyad, G. Piatetsky- Shapiro, P. Smyth, and R. Uthurusamy, 1–30. Menlo Park, Calif.: AAAI Press.

[GUANLIN ET AL, 2015]_An Intelligent Analysis and Mining System for Urban Lighting Information

[GOLDING & ROSENBLOOM, 1991]_-A Supervisor Agent for Urban Traffic Monitoring

-improuving rule-based system

[GEORGE, 2011] GEORGE, Lars. HBase: the definitive guide. " O'Reilly Media, Inc.", 2011.

[HOUDESHEL & WATSON, 1987] _Houdeshel, G. and H. Watson, "The Management Information and Decision Support (MIDS) System at Lockheed-Georgia", MIS Quarterly, 11, 1, March 1987, 127-140

[HENDERSON, 97]Henderson JA (1997) *Community Planning And Development.* Lecture notes Spring Semester, 1997, Whittier College, California,

[HBASE, 2015] TEAM, Apache HBase. Apache HBase Reference Guide. 2015.

[HUGGONARD & ROY, 82] HUGGONARD, J. C. AND B. Roy, "Classement des prolongements des lignes de metro en banlieue parisienne:Presentation d'une methode multicritere originale," Compte Rendus des 14e journees du groupe de travail europeen sur l'aide a la Decision mulitcritere, European Institute for Advanced Studies in Management, Brussels, 1982.

[HAAGSMA & JOHANNS, 94] HAAGSMA, I. G. et JOHANNS, R. D. Decision support systems: An integrated and distributed approach. COMPUTATIONAL MECHANICS, INC., BILLERICA, MA 01821(USA)., 1994, p. 205-212.

[HAYES, 81] Hayes JR (1981) The Complete Problem Solver. Philadelphia Franklin Institute Press.

[HADOOP, 2016] HADOOP, Apache. Welcome to Apache Hadoop!(2016).

[HAETTENSCHWILER, 01] HAETTENSCHWILER, Plus. Neues anwenderfreundliches konzept der entscheidungsunterstützung. Zurich, vdf Hochschulverlag AG, 2001, p. 189-208..

[HALL ET AL, 09]_ HALL, Mark, FRANK, Eibe, HOLMES, Geoffrey, et al. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 2009, vol. 11, no 1, p. 10-18.

[INMON, 05]_INMON, William H. Building the data warehouse. John wiley & sons, 2005.

[JAKŠIĆ & RAKOČEVIĆ, 2012] JAKŠIĆ, Maja Levi et RAKOČEVIĆ, Slađana Barjaktarović. Proceedings of the XIII International Symposium SymOrg 2012: Innovative Management and Business Performance. FON, 2012.

[KONIDARI & MAVRAKIS, 07] Konidari, P., and Mavrakis, D. (2007). A multi-criteria evaluation method for climate change mitigation policy instruments. *Energy Policy*, 35(12): 6235-6257.

[KAKLAUSKAS, 2015] KAKLAUSKAS, Arturas. Biometric and intelligent decision making support. Springer, 2015.

[**KISILEVICH ET AL, 2013**] Kisilevich S, Keim D, Rokach L (2013) A GIS-based decision support system for hotel room rate estimation and temporal price prediction: The hotel brokers' context. Decision Support Syst54(2):1119–1133

[**KEEN & MORTON, 78**] Keen, P.G.W., and Morton, M.S. Decision Support Systems: an Organizational Perspective Addison-Wesley, Reading, MA, 1978.

[**KEENEY, 92**] Keeney, R. L. Value-Focused Thinking: A Path to Creative Decision

Making. Harvard University Press, Cambridge, Massachusetts.

[**KULKARNI ET AL, 07**] Kulkarni, U., Power, D. J., & Sharda, R. (Eds.). (2007). Decision support for global enterprises (Vol. 2). Springer Science & Business Media.

[**KAR, 2015**] KAR, Arpan Kumar. A hybrid group decision support system for supplier selection using analytic hierarchy process, fuzzy set theory and neural network. Journal of Computational Science, 2015, vol. 6, p. 23-33.

[**KELLER,02**] KELLER, Frank. Naive Bayes Classifiers. Connect. Stat. Lang. Process. Course Univ. Saarlandes, 2002.

[**LAURINI, 02**] LAURINI, Robert. Information systems for urban planning: a hypermedia cooperative approach. CRC Press, 2002.

[**LAURINI & THOMPSON, 92**] LAURINI, Robert et THOMPSON, Derek. Fundamentals of spatial information systems. Academic press, 1992.

[**LEUNG¸07**] LEUNG, K. Ming. Naive bayesian classifier. Polytechnic University Department of Computer Science/Finance and Risk Engineering, 2007.

[**LI ET AL, 2012**] LI, Xiaolong, PAN, Gang, WU, Zhaohui, et al. Prediction of urban human mobility using large-scale taxi traces and its applications. Frontiers of Computer Science, 2012, vol. 6, no 1, p. 111-121.

[**MCCARTHY ET AL, 55**] McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (1955). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence". Archived from the original on 26 August 2007. Retrieved 30 August 2007.

[**MCCARTHY, 07**] McCarthy, John (12 November 2007). "What Is Artificial Intelligence?" http://www-formal.stanford.edu/jmc/whatisai/whatisai.html

[MALCZEWSKI, 99] J. Malczewski, GIS and Multicriteria Decision Analysis. John Wiley & Sons, New York, 1999.

[MARINONI, 05] MARINONI, Oswald. A stochastic spatial decision support system based on PROMETHEE. International Journal of Geographical Information Science, 2005, vol. 19, no 1, p. 51-68.

[MCKENNEY & KEEN, 74] McKenney, J., and Keen, P. How Managers' Minds Work. Harvard Business Review (52:3) 1974, pp 79-90.

[METTAN, 92] S.Mettan la négociation : son rôle sa place dans l'aménagement du territoire et la protection de l'environnement Paris 1992.

[MARAKAS, 99] Marakas, G. M. (1999). Decision support systems in the twenty-first century. Upper Saddle River, N.J., Prentice Hall.

[MERRILL, 2012] MERRILL, Ray M. Fundamentals of Epidemiology and Biostatistics. Jones & Bartlett Publishers, 2012.

[MAY ET AL, 96] May AD, Mitchell G and Kapiszewska D (1996) 'The Quantifiable City: the Development of a Modelling Framework for Urban Sustainability Research'. In *Information Systems and Processes for Urban Civil Engineering Applications*, edited by Ugo Schiavoni, Rome, November 21–22, 1996. Published by the European Communities in 1998, European Commission EUR 18325 EN, ISBN 92–8283734–3. pp 124–42.

[MITCHELL, 97] T.M.. Mitchell, *Machine Learning*. Singapore: McGraw-Hill Companies Inc., 1997.

[O'LEARY, 2013]O'LEARY, Daniel E. Artificial intelligence and big data. *IEEE Intelligent Systems*, 2013, no 2, p. 96-99.

[OFLI ET AL, 2016] OFLI, Ferda, MEIER, Patrick, IMRAN, Muhammad, et al. Combining Human Computing and Machine Learning to Make Sense of Big (Aerial) Data for Disaster Response. Big Data, 2016.

[POWER, 07] Power, D.J. A Brief History of Decision Support Systems. DSSResources.COM, World Wide Web, ttp://DSSResources.COM/history/dsshistory.html, version 4.0, March 10, 2007.

[Power, 02] Power, D. J. (2002). Decision support systems : concepts and resources for managers. Westport,Conn., Quorum Books.

[PHIT et AL, 2010] PHITHAKKITNUKOON, Santi, VELOSO, Marco, BENTO, Carlos, et al. Taxi-aware map: Identifying and predicting vacant taxis in the city. In : Ambient Intelligence. Springer Berlin Heidelberg, 2010. p. 86-95.

[PARK & STENSTROM, 08] PARK, Mi-Hyun et STENSTROM, Michael K. Classifying environmentally significant urban land uses with satellite imagery. Journal of environmental management, 2008, vol. 86, no 1, p. 181-192.

[Pingdom, 2011] Pingdom, “Internet 2011 in Numbers,” tech. blog, 17 Jan. 2012; http://royal.pingdom.com/2012/01/17/internet- 2011-in-numbers.

[Rockart, 1979] Rockart, J. F. "Chief Executives Define Their Own Data Needs," Harvard Business Review, 67, 2 March-April 1979, 81-93.

[Roy, 96]_B. Roy. Multicriteria Methodology for Decision Aiding, volume 12 of Nonconvex Optimization and its Applications. Kluwer Academic Publishers, Dordrecht, 1996.

[Roy, 05] Roy, B. (2005). Paradigms and challenges. In *Multiple criteria decision analysis: State of the art surveys* (pp. 3-24). Springer New York.

[Robbins, 02] Robbins, S.P. Organizational Behavior Prentice Hall of India, New Delhi, 2002.

[Romeo, 2015] _Romeo, Leonardo. 2015. What is territorial development? GREAT insights Magazine, Volume 4, Issue 4. June/July 2015.

[Simon, 77]_H.A. Simon, «The new science of management decision».Prentice –Hall, Administration et processus de décision 3 mes édition, Paris,1977.

[Saracoglu, 2013]_ Saracoglu, B.O. (2013). "Selecting industrial investment locations in master plans of countries". European J. of Industrial Engineering (Inderscience Enterprises Ltd.) 7 (4): 416–441.

[Sayad, 2016] An Introduction to Data Mining http://www.saedsayad.com.

[Saaty a, 08] Saaty, Thomas L.; Peniwati, Kirti (2008). Group Decision Making: Drawing out and Reconciling Differences. Pittsburgh, Pennsylvania: RWS Publications.

[SAATY B, 08] Saaty, Thomas L. (June 2008). "Relative Measurement and its Generalization in Decision Making: Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors – The Analytic

[SUMATHI & SIVANANDAM, 06] SUMATHI, Sai et SIVANANDAM, S. N. Introduction to data mining and its applications. Springer, 2006.

[SPRAGUE & CARLSON, 82] Sprague, R. H. and E. D. Carlson (1982). Building effective decision support systems. Englewood Cliffs, N.J., Prentice-Hall.

[SPATZ ET AL, 2014] Joachim Spatz & Stefan Schaal Max Planck Intelligent Systems Research Institute for Intelligent Systems 2014

[TWEEDALE ET AL, 2016] TWEEDALE, Jeffrey W., NEVES-SILVA, Rui, JAIN, Lakhmi C., et al.Intelligent Decision Technology Support in Practice. Springer Publishing Company, Incorporated, 2015.

[VELASQUEZ & HESTER, 2013] VELASQUEZ, Mark et HESTER, Patrick T. An analysis of multi-criteria decision making methods. *International Journal of Operations Research*, 2013, vol. 10, no 2, p. 56-66.

[VILLEDURABLE, 2014] web site for urban information and offer solution for urban problem villedurable.org

[VMWARE, 2016] https://www.vmware.com/products/workstation/features.

_[WANG, 09] Wang H-F, Wu C-T (2009) A mathematical model for product selection strategies in a recommender system. Expert Syst Appl 36(3):7299–7308

[WCED, 87] WCED (World Commission on Environment and Development) 1987. Our Common Future: Report of the World Commission on Environment and Development. Oxford: Oxford University Press.

[WEEKS, 2010] WEEKS, John R. Defining urban areas. In: Remote Sensing of Urban and suburban Areas. Springer Netherlands, 2010. p. 33-45.

[WEEKS, 08] Weeks JR (2008) Population: an introduction to concepts and issues. 10th edn. Wadsworth Thomson Learning, Belmont, CA Hierarchy/Network Process". Review of the Royal Academy of Exact, Physical and Natural Sciences, Series A: Mathematics (RACSAM) 102 (2): 251–318.

[WITTEN & FRANK, 05] WITTEN, Ian H. et FRANK, Eibe. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.

[WARRENS ET AL, 08] WARRENS, Matthijs Joost, et al. Similarity coefficients for binary data: properties of coefficients, coefficient matrices, multi-way metrics and multivariate coefficients. Psychometrics and Research Methodology Group, Leiden University Institute for Psychological Research, Faculty of Social Sciences, Leiden University, 2008.

[WAN ET AL, 90] WANG, Fangju, BRENT HALL, G., et SUBARYONO. Fuzzy information representation and processing in conventional GIS software: database design and application. International Journal of Geographical Information System, 1990, vol. 4, no 3, p. 261-283.Impoving rule based systems An Intelligent Analysis and Mining System for Urban Lighting Information.

[YOU ET AL, 2007] LE DIPLOME, DE MAGISTER. Proposition d'un modèle Décisionnel en Aménagement du Territoire par Utilisation des SIG et des Réseaux de Neurones.

[YANG ET AL, 2015] YANG, Yang, TANG, Jingyin, LUO, Hao, et al. Hotel location evaluation: a combination of machine learning tools and web GIS. International Journal of Hospitality Management, 2015, vol. 47, p. 14-24.

[ZHANG ET AL, 2010] ZHANG, Chen, DE STERCK, Hans, ABOULNAGA, Ashraf, et al. Case study of scientific data processing on a cloud using hadoop. In: High performance computing systems and applications. Springer Berlin Heidelberg, 2010. p. 400-415.