

UNA GUIDA PRATICA PER LO SVILUPPO DEI PROGETTI AI NELLE PMI



A cura di Damiano Cosma De Angelis

Sommario

Dalla crescita lineare alla discontinuità
Grandi imprese e PMI davanti alla sfida dell'AI
L'intelligenza esperienziale e il metodo
Il processo standard per i progetti AI
Comprensione del Business
Comprensione dei Dati
Preparazione dei Dati
Modellizzazione
Valutazione
Implementazione
Siamo in onda: come procedere?
Conclusioni

Negli ultimi decenni una enorme quantità di ricercatori si sono dedicati ad un progetto antico: lo sviluppo di algoritmi che imparano, imitando gli esseri umani. Il ritmo è stato talmente rapido da essere incompatibile con i tempi di assimilazione del tessuto industriale e formativo. L'articolo presenta il processo standard CRISP-DM, riletto ed arricchito grazie ad esperienze reali acquisite nella implementazione di progetti AI in ambito industriale. Una guida semplice e di immediata adozione per le PMI.

Dalla crescita lineare alla discontinuità

Uno dei tratti distintivi della contemporaneità è lo sviluppo tecnologico, che disegna la traiettoria della evoluzione delle persone, e delle aziende. Nei decenni questo sviluppo ha costantemente aumentato la sua velocità, in una condizione di accelerazione continua.

L'Intelligenza Artificiale introduce in questo senso una novità assoluta. Si tratta di una perturbazione nella traiettoria "normale" dello sviluppo, di un salto, ma un salto talmente grande, che mette in oggettiva difficoltà noi tutti, e cambia in qualche modo le regole. È per questo motivo che affrontare il tema della Intelligenza Artificiale, è diventata una sfida per chiunque. Soprattutto per le aziende. Le tecniche di Intelligenza Artificiale aprono prospettive - su diverse dimensioni - che consentono miglioramenti che nessuno può ignorare. Si pensi ad un programmatore: l'utilizzo - professionale - della AI porta una moltiplicazione di produttività, come usa dire, per 10. Un programmatore esperto, capace di utilizzare correttamente gli strumenti della IA, produce in un giorno ciò che prima produceva in 10 giorni. Inoltre - incredibilmente - il prodotto finale è complessivamente molto migliore di quello sviluppato in modo tradizionale.

Non adottare queste tecniche non è quindi più una scelta ideologica, ma una questione di sopravvivenza professionale.

Grandi imprese e PMI davanti alla sfida dell'AI

Per le aziende vale lo stesso concetto, ma le aziende non sono tutte uguali. Usando una tassonomia ampiamente condivisa possiamo collocare le aziende in due ambiti principali. Quelle che vivono principalmente dell'esperienza e del talento delle persone che le compongono, altre che traggono la loro forza da processi strutturati ed organizzazioni complesse.

Nelle prime sono collocate le piccole medie aziende Italiane, mentre nelle seconde possiamo collocare le grandi imprese - spesso internazionali. Queste due tipologie di aziende affrontano, in genere, in modo diverso la sfida evolutiva. Le grandi aziende hanno le risorse per cercare soluzioni al di fuori di sé: un nuovo fornitore, una nuova tecnologia, un nuovo servizio. Le PMI cercano di orga-

nizzarsi con le proprie forze, cercano di "imparare" il nuovo, per farlo diventare un nuovo elemento di valore dell'azienda.

Questo ultimo modello di sviluppo presenta però un vincolo essenziale: le risorse interne devono avere le capacità tecniche per affrontare il nuovo. Solitamente, quando c'è da adeguare l'azienda alle mutate condizioni tecnologiche, la parte "tecnica" delle PMI viene chiamata a svolgere quel ruolo. Caso tipico è il settore IT, al quale vengono di default assegnati questi obiettivi. E spesso è proprio il settore IT a cui si chiede di affrontare i progetti AI. Ma con l'Intelligenza Artificiale le cose cambiano sensibilmente, sia nel metodo che negli obiettivi.

La grande azienda può continuare nella sua modalità, accedendo ad ecosistemi di risorse esterne. Le tecnologie della AI sono talmente recenti e talmente veloci nella loro evoluzione, che le risorse interne nelle PMI possono non avere gli strumenti sufficienti ad affrontare la sfida.

Però sono proprio le PMI che possono trarre dalle tecnologie AI il massimo vantaggio in termini di evoluzione. Scopo della Intelligenza Artificiale è, come noto, quello di aumentare, amplificare le capacità della intelligenza umana, intesa non come intelligenza "teorica", bensì come capacità euristica derivante dall'esperienza. Potremmo definirla "Intelligenza esperienziale", l'intelligenza che risolve problemi, perché ha imparato dall'esperienza. Ma questa è l'intelligenza che contraddistingue le PMI e ne costituisce una componente di valore fondamentale. Allora l'intelligenza artificiale per le PMI assume questa connotazione: la vera risorsa da valorizzare è quella che l'impresa già possiede: la sua conoscenza, la sua esperienza, i suoi dati.

L'intelligenza esperienziale e il metodo

Anche dal punto di vista dei casi d'uso, la AI si pone in modo diverso nei due ambiti aziendali su delineati. Nelle realtà fortemente organizzate, la AI favorisce l'efficienza e la standardizzazione dei processi, proprio perché è lì concentrata l'esperienza della grande azienda. Nelle PMI, invece, la AI trova il suo terreno più fertile nelle competenze e nell'esperienza operativa delle persone, traducendole in dati, modelli e processi automatici.

È in questo insieme di conoscenza umana e capacità tecnica, che si gioca il vero valore dell'AI per il tessuto produttivo delle PMI. Una caratteristica interessante dei progetti AI è che richiedono meno competenze tecniche di quanto si pensi. Il problema non sono gli strumenti; ad esempio gli algoritmi di Machine Learning sono pubblici, disponibili liberamente per chiunque li sappia utilizzare. Il problema è avere uno schema da seguire, una guida, per realizzare i progetti che utilizzano AI, a cui oggi non siamo ancora abituati.

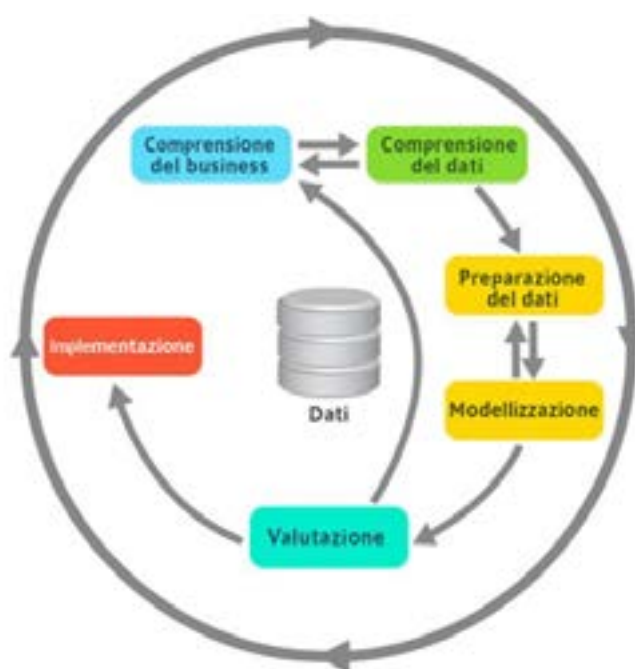
In altre parti del mondo quelle che noi vediamo come novità assolute erano già all'attenzione delle università e delle aziende più avanzate. Già alla fine del secolo scorso è stata proposta una metodologia chiamata CRISP-DM (CRoss-Industry Standard Process for Data Mining) che costituiva, e costituisce ancor più oggi se adeguatamente aggiornata, una guida preziosa per chi voglia approcciare i progetti AI.

Si consideri che il Data Mining, lo "scavare nei dati", è alla base della costruzione di modelli di Intelligenza Artificiale, di cui il Machine Learning costituisce la tecnologia operativa. L'attualità di metodo qui proposto, sta nel fatto di offrire una cornice essenziale, estremamente utile per le PMI che possono trarre, avendo tracciata una strada chiara e ben standardizzata, le indicazioni per affrontare i primi progetti di Intelligenza Artificiale. Il processo è stato utilizzato più volte, nella azione di supporto ad aziende che, dopo diversi tentativi, hanno dovuto prendere atto che la delega di tali progetti, per sua natura, non è possibile. Se è vero che la A di AI significa "Aumentata", più che "Artificiale", non possiamo delegare l'aumento della nostra intelligenza - sempre nei termini su delineati - ad altri.

È l'intelligenza delle nostre PMI, delle persone che le compongono, che dobbiamo trovare il modo di "Aumentare".

Il processo standard per i progetti AI

Nella figura è rappresentato il processo standard, che vede coinvolte diverse figure aziendali e, facendo riferimento alle PMI, nessuno è escluso a priori dal partecipare al processo. Una sottolineatura necessaria è sulle frecce che uniscono le



varie fasi. Significano che il processo prevede dei passi in avanti, ma anche passi indietro verso la fase, o le fasi, precedenti.

Comprensione del Business

Qualsiasi iniziativa di impresa richiede una fase come questa. Si vuole valutare la possibilità di creare qualcosa che non c'è. In questa fase, nelle PMI, le figure essenziali sono quelli che posseggono il know-how che caratterizza l'azienda.

I fondatori, i soci operativi, coloro che fanno e che hanno tesaurizzato l'esperienza.

Non è realisticamente possibile affrontare questa fase senza prima avere imparato le basi, i fondamentali, della AI.

L'esperienza mostra che la strada più semplice è quella di una formazione di base, rivolta proprio alle figure di riferimento dell'azienda. Non è questa una fase tecnica, da delegare a strutture "specialistiche" interne, o esterne. È la fase in cui si "inventa" il nuovo, grazie all'esperienza stratificata nel tempo.

Una azione indispensabile iniziale è dare un titolo al progetto. Un titolo chiaro, breve, univoco. L'esperienza ci insegna che il titolo è un elemento imprescindibile per il successo di qualsiasi progetto. Se non riusciamo a dare un titolo, non possiamo andare avanti.

L'obiettivo primario della analisi è la comprensione profonda di cosa si vuole ottenere. Si risponde a "cosa si vuole realizzare?", o anche "quale è l'idea che abbiamo in mente?".

Potremmo scoprire di avere in mente obiettivi diversi, a volte contrastanti. Allora bisogna chiarire gli aspetti fondamentali: stiamo parlando di AI predittiva o generativa? Quali aspetti possono pregiudicare il risultato? Quali risorse siamo disposti a mettere in campo? Se non riusciamo a rispondere, non possiamo proseguire. È estremamente raccomandato dare ugualmente una stima di quali risorse, materiali ed umane, pensiamo di poter allocare per il progetto. Una componente fondamentale da prevedere in questa fase: quali risultati/vantaggi mi aspetto. Come faccio a valutarli una volta che il processo è pronto?

Siamo un'un'area dove il componente più prezioso è il buonsenso, la capacità tutta umana di scegliere in quale modo capire se un processo - pur complesso - fa quello che "deve" fare oppure no.

Comprensione dei Dati

Arriviamo alla seconda fase se, e solo se, abbiamo risposto alle domande della fase iniziale.

Questa è la fase che dà il nome al metodo, ovvero "Data Mining". Abbiamo i dati che contengono le informazioni che pensiamo possano servirci? Se vi sono dati informatici non digitali, provvediamo a digitalizzarli. Ovvero abbiamo dati che forse contengono ciò che cerchiamo? Se la risposta è sì, in questa fase i dati - digitali o meno - vengono raccolti, e se ne descrivono le proprietà fondamentali. Siamo in presenza di Big Data? E se sì, come impattano il nostro progetto, con quale piattaforma li gestiamo, disponiamo di algoritmi di Machine Learning opportuni?

Questi dati vengono esplorati con le tecniche che vanno sotto il nome comune di "EDA" (Explorative Data Analysis). Si realizzano grafici, si fanno analisi statistiche, e se si mostrano elementi - per qualsiasi motivo - potenzialmente problematici, si torna indietro alla fase di "Comprensione del Business", ricordando che il ritorno indietro non è un errore, è parte del processo.

Se invece pensiamo di essere pronti, procediamo verso la fase successiva.

A questa fase partecipano sia i portatori di conoscenza aziendale, sia gli esperti di Analisi dei Dati, in una sorta di staffetta dal know-how alla tecnologia dei dati.

Preparazione dei Dati

Qui si preparano i dati, ormai tutti digitali, per la successiva fase di modellazione. I dati vengono generalmente trattati, sia per diminuire il rumore sia per arricchirli ancora. L'arricchimento può venire da dati esterni, che richiedono l'applicazione di tecniche di integrazione e di riconciliazione. Quando le fonti sono diverse è normale che vi siano difformità formali, oppure duplicazioni, che emergono solo dopo aver proceduto a trasformazioni, che possono comportare uno sforzo programmatico, computazionale e temporale sensibile.

Infine si organizzano i dati, così come gli algoritmi che si ipotizza di utilizzare, si aspettano e si opera una selezione dei dati, con lo scopo di eliminare, per quanto possibile, il rumore.

Per rimanere nella terminologia classica, ci troviamo nella fase di "FE" (Feature Engineering).

Le competenze sono quelle di ingegneria dei dati e di esperti di Machine Learning.

Modellizzazione

Tutto è pronto per applicare gli algoritmi opportuni ai dati che sono stati preparati nella fase precedente. Si tratta di una fase certamente tecnica, vengono definiti i criteri di validazione dei modelli proposti in modo da selezionare quello "giusto". Vengono quindi applicati tutti gli algoritmi compatibili con i dati e con il caso d'uso. Può essere qui interessante valutare l'utilizzo delle tecniche di AutoML, che automatizzano questo processo in genere piuttosto lungo.

In questo momento viene definita l'accettabilità del modello, in termini di criteri di valutazione numerica del risultato dei modelli. Mandatoria la stesura di un report di dettaglio sui modelli candidabili e su quello selezionato. Il report può essere di grande aiuto nel momento in cui si debba cambiare qualcosa.

Ed a questo proposito, questa fase è "Naturalmente" iterativa rispetto alla precedente, nel senso che se non si riesce ad ottenere modelli soddisfacenti

- in senso tecnico - si deve chiedere a chi precede di ripercorrere la fase di *"Preparazione dei dati"*. Le figure coinvolte sono esperti di Machine Learning. Ci si allontana dalla necessità di conoscenza del business, ci si inoltra nella fase di tecnica degli algoritmi.

Valutazione

Sembra una ripetizione, ma è invece la fase di valutazione non tecnica, ma di business. Arrivati a questo punto, con il modello selezionato, raggiungiamo gli obiettivi di business? Il gruppo che si occupa di questa fase è un gruppo diverso dai precedenti. È costituito da risorse che devono *"usare"* il sistema, alimentandolo con dati diversi da quelli utilizzati nelle fasi precedenti.

Se devo interrogare uno studente per valutare se ha *"capito"* ciò che ha studiato, gli farò domande che hanno forma diversa dal testo su cui ha imparato. Lo stesso per i modelli: vanno testati con datasets di *"Test"* che non devono essere derivati dai dati precedenti.

Questa condizione, non banale né a volte semplice da realizzare, e va prevista già nella primissima fase iniziale.

Non solo vanno usati i modelli con dati nuovi, ma qui è necessario anche confrontare il risultato tecnico da quello di business. In altre parole: il modello risponde bene dal punto di vista delle metriche di qualità, ma il risultato finale, quello che arriva al *"cliente"*, interno o esterno che sia, è quello che volevamo?

Non è raro che solo in questo momento ci si accorga che qualcosa non va, al punto da scoprire di dover - vedi freccia nella figura - tornare alla fase iniziale di *"Comprensione del Business"*.

Non si deve considerare una tale eventualità come un errore, ma come una iterazione la cui necessità si rende palese solo in questo momento. Se invece tutto è come prevedevamo, si passa alla fase di *"Implementazione"* o - in termini più generali - di *"Esercizio"*.

Implementazione

L'intelligenza Artificiale può essere sviluppata su piattaforme diverse, che possono essere molto dissimili fra loro.

Tutte le aziende hanno la loro infrastruttura IT e soprattutto le relative competenze interne. Conciliare questi due ambiti è di importanza fondamentale. La scelta della piattaforma di esercizio influenza quindi quella di sviluppo, e viceversa.

Si tratta pur sempre di software, ma la scelta della piattaforma necessaria per mettere in esercizio i modelli di Machine Learning sviluppati, potrebbe non essere allineata alle tecnologie consuete in azienda.

Questo aspetto ha un impatto talmente rilevante, che va trattato nella fase iniziale. In altri termini, potrei dover scegliere alcune tecnologie di Machine Learning, piuttosto che altre in funzione delle piattaforme - e delle competenze - interne di chi dovrà poi mettere in esercizio il sistema.

Una volta soddisfatti i requisiti, si passa alla verifica della fruibilità del sistema da parte del *"cliente"* - interno o esterno - finale, si preparano manuali utente e guide operative per la gestione della infrastruttura.

Va valutato se inserire la componente AI nel sistema di monitoraggio ICT, poiché entra nella catena applicativa un oggetto, come il modello AI, che non è ancora una parte *"matura"* della infrastruttura ICT.

Siamo in onda: come procedere ?

I sistemi basati su AI sono per definizione dinamici. Possono variare gli obiettivi di business, possono cambiare i dati disponibili, possono arrivare di nuovi che ci fanno decidere di sfruttarne le potenzialità informative, o meno. Questa prospettiva - non obbligatoria - è indicata dalla freccia circolare circonda lo schema.

Ma questa è una ulteriore novità dei progetti AI che, essendo legati a dati e condizioni mutevoli, possono richiedere questa sorta di iterazione continua nel tempo.

Conclusione

Un processo come quello descritto, non è solo una procedura, ma vuole offrire una visione del mondo. In questo caso è un mondo nuovo e quindi, inevitabilmente, richiede uno sforzo, un allontanamento dalla consuetudine. La procedura proposta, il CRISP-DM rivisitato, può supportare



(*) **Damiano Cosma De Angelis**

Ingegnere, consulente e formatore AI.

le aziende, in particolare le PMI, ad iniziare questo percorso nuovo, ma di grandissima potenzialità.

Keywords: *Intelligenza Artificiale (AI); PMI; CRISP-DM; Data Mining; Machine Learning; Processo standard AI; Intelligenza esperienziale; Comprensione del Business; Comprensione dei Dati; Preparazione dei Dati; Modellizzazione; Valutazione e Implementazione; Ecosistema aziendale; Automazione dei processi; Innovazione nelle PMI*

A PRACTICAL GUIDE TO DEVELOPING AI PROJECTS IN SMES

Over the past decades, a vast number of researchers have devoted themselves to an age-old endeavor: the development of algorithms that can learn by imitating human beings. The pace of this progress has been so rapid that it has become incompatible with the rate at which the industrial and educational fabric can absorb it.

This article presents the standard CRISP-DM process, reinterpreted and enriched by real-world experience gained through the implementation of AI projects in industrial environments. It is designed as a simple and immediately adoptable guide for small and medium-sized enterprises.

By Damiano Cosma De Angelis

Summary

From Linear Growth to Discontinuity
Large Enterprises and SMEs Facing the AI Challenge
Experiential Intelligence and the Method
The Standard Process for AI Projects
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Implementation
We're On Air: What's Next?
Conclusions

From Linear Growth to Discontinuity

One of the defining features of our contemporary age is technological development, which shapes the trajectory of evolution for both individuals and companies. Over the decades, this development has continuously increased its speed, creating a condition of perpetual acceleration. Artificial Intelligence introduces, in this regard, an entirely new phenomenon. It represents a

disturbance in the “normal” trajectory of development—a leap forward, and such a large leap that it objectively challenges us all and, in many ways, changes the rules of the game.

For this reason, tackling the topic of Artificial Intelligence has become a challenge for everyone—especially for businesses.

AI techniques open up new perspectives—across multiple dimensions—that enable improvements no one can afford to ignore. Consider a software developer: the professional use of AI can multiply productivity by a factor of ten. An experienced programmer capable of correctly using AI tools can produce in a single day what previously required ten. Moreover—remarkably—the final product is often of higher overall quality than one developed through traditional methods.

Therefore, not adopting these techniques is no longer an ideological choice; it has become a matter of professional survival.

Large Enterprises and SMEs Facing the AI Challenge

The same principle applies to companies, but not all companies are the same. Using a widely accepted taxonomy, we can distinguish two main types of organizations:

those that rely primarily on the experience and talent of their people, and those that draw their strength from structured processes and complex organizations.

The first category includes small and medium-sized Italian enterprises, while the second encompasses large—often international—corporations.

These two types of companies usually approach the evolutionary challenge in very different ways. Large enterprises have the resources to seek solutions outside themselves: a new supplier, a new technology, a new service. SMEs, on the other hand, tend to organize with their own means, striving to “learn” the new so as to turn it into a new source of value for the company.

However, this second development model presents an essential constraint: internal resources must possess the technical capabilities needed to face change.

Usually, when the company must adapt to new

technological conditions, the “technical” part of the organization—typically the **IT department**—is called upon to take on this role. A classic case is when the IT department is automatically assigned responsibility for AI projects.

But with Artificial Intelligence, things change significantly—both in method and in objectives.

A large company can continue operating in its traditional mode, accessing external ecosystems of expertise and service providers.

For SMEs, however, the situation is different: AI technologies are so recent and evolve so rapidly that internal resources may lack the necessary tools to take on the challenge effectively.

Yet, paradoxically, it is precisely **SMEs** that stand to gain the greatest advantage from AI technologies in terms of evolution and growth.

The purpose of Artificial Intelligence, as is well known, is to enhance and amplify human intelligence—not in its “theoretical” sense, but in its **heuristic capacity**, the ability to solve problems through experience.

We could call this “**experiential intelligence**” — the kind of intelligence that learns by doing.

And this is exactly the type of intelligence that distinguishes SMEs and forms a fundamental component of their value.

Therefore, for SMEs, Artificial Intelligence takes on a specific connotation: the real resource to be enhanced is what the enterprise already possesses—its **knowledge, experience, and data**.

Experiential Intelligence and the Method

Even from the point of view of use cases, AI manifests differently within the two types of organizations described above.

In highly structured companies, AI primarily promotes efficiency and process standardization—because that is where the organization’s experience resides.

In SMEs, however, AI finds its most fertile ground in the **skills and operational experience of people**, translating this human know-how into data, models, and automated processes.

It is within this combination of **human knowledge and technical capability** that the real value of AI for the SME production fabric lies.

An interesting characteristic of AI projects is that they require **fewer technical skills than one might think**.

The issue is not the availability of tools-Machine Learning algorithms, for instance, are public and freely accessible to anyone who knows how to use them.

The real issue is having a **framework**, a **guide**, to follow when implementing AI projects-something that most organizations are not yet accustomed to. In other parts of the world, what we now consider groundbreaking innovations were already being studied by universities and leading companies years ago.

As early as the end of the last century, a methodology called **CRISP-DM** (*Cross-Industry Standard Process for Data Mining*) was proposed.

It represented-and still represents today, when properly updated-a valuable guide for anyone wishing to approach AI projects methodically.

It should be remembered that **Data Mining**, or "*digging into data*," forms the basis of constructing Artificial Intelligence models, with **Machine Learning** serving as the operative technology that brings them to life.

The relevance of the CRISP-DM method lies precisely in providing an **essential framework**, extremely useful for SMEs, offering them a **clear and standardized roadmap** for tackling their first AI projects.

This process has been applied many times to support companies that, after several unsuccessful attempts, had to recognize a key fact: **delegating such projects is inherently impossible**.

If the "A" in AI stands for **Augmented** rather than **Artificial**, then we cannot delegate the augmentation of our intelligence-at least in the sense described above-to someone else.

It is the intelligence of our SMEs, and of the people who make them up, that we must find a way to **augment**.

The standard process for AI projects

The figure represents the standard process, which involves different company figures and, referring to SMEs, no one is excluded a priori from participating in the process. A necessary emphasis is

on the arrows that join the various phases. They mean that the process involves steps forward, but also steps back towards the previous phase, or phases.

Understanding the Business

Any business initiative requires a phase like this. You want to evaluate the possibility of creating something that doesn't exist. At this stage, in SMEs, the essential figures are those who possess the know-how that characterizes the company. The founders, the operating partners, those who know and who have hoarded the experience. It is not realistically possible to face this phase without first having learned the basics, the fundamentals, of AI.

Experience shows that the easiest way is that of basic training, aimed precisely at the company's reference figures. This is not a technical phase, to be delegated to internal or external "*specialist*" structures. It is the phase in which the new is "*invented*", thanks to experience stratified over time. An initial indispensable action is to give a title to the project. A clear, short, unambiguous title. Experience teaches us that the title is an essential element for the success of any project. If we can't give a title, we can't move forward.

The primary objective of the analysis is a deep understanding of what you want to achieve. The answer is "*what do you want to achieve?*", or even "*what is the idea we have in mind?*".

We may find that we have different, sometimes conflicting, goals in mind. So we need to clarify the fundamental aspects: are we talking about predictive or generative AI? What aspects can affect the result? What resources are we willing to put in place? If we cannot answer, we cannot continue. It is highly recommended that you also give an estimate of what resources, material and human, we think we can allocate to the project. A key component to predict at this stage: what outcomes/benefits do I expect How do I evaluate them once the process is ready?

We are an area where the most valuable component is common sense, the all-human ability to choose how to understand whether a process - however complex - does what it "*must*" do or not.

Understanding Data

We arrive at the second phase if, and only if, we have answered the questions of the initial phase. This is the phase that gives the method its name, i.e. *"Data Mining"*. Do we have the data that contains the information we think we might need? If there is non-digital information, we digitize it. That is, do we have data that perhaps contains what we are looking for? If the answer is yes, in this phase the data - digital or not - are collected, and their fundamental properties are described. Are we in the presence of Big Data? And if so, how do they impact our project, with which platform do we manage them, do we have appropriate Machine Learning algorithms?

This data is explored with techniques that go by the common name of *"EDA"* (Explorative Data Analysis). Graphs are made, statistical analyses are made, and if you show elements - for any reason - potentially problematic, you go back to the *"Understanding the Business"* phase, remembering that going back is not a mistake, it is part of the process.

If, on the other hand, we think we are ready, we proceed to the next phase.

Both bearers of corporate knowledge and Data Analysis experts participate in this phase, in a sort of relay race from know-how to data technology.

Data Preparation

Here the data, now all digital, are prepared for the next modeling phase. Data is generally processed, both to reduce noise and to further enrich it. Enrichment can come from external data, which requires the application of integration and reconciliation techniques. When the sources are different, it is normal for there to be formal discrepancies, or duplications, which emerge only after having proceeded with transformations, which can involve a significant programmatic, computational and temporal effort.

Finally, the data are organized, as well as the algorithms that are hypothesized to be used, a selection of the data is expected and carried out, with the aim of eliminating, as far as possible, the noise.

To stay in the classic terminology, we are in the *"FE"* Feature Engineering phase).

The skills are those of data engineering and Machine Learning experts.

Modeling

Everything is ready to apply the appropriate algorithms to the data that has been prepared in the previous phase. This is certainly a technical phase, the validation criteria of the proposed models are defined in order to select the *"right"* one. All algorithms that are compatible with the data and the use case are then applied. It may be interesting here to evaluate the use of AutoML techniques, which automate this generally rather long process. At this moment the acceptability of the model is defined, in terms of the criteria for numerical evaluation of the result of the models. The drafting of a detailed report on the eligible models and on the selected one is mandatory. The report can be of great help when you need to change something.

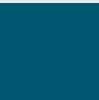
And in this regard, this phase is *"Naturally"* iterative compared to the previous one, in the sense that if you cannot obtain satisfactory models - in a technical sense - you have to ask those who precede to retrace the *"Data Preparation"* phase. The figures involved are Machine experts

Learning. We move away from the need for business knowledge, we enter the technical phase of algorithms.

Evaluation

It seems like a repetition, but it is instead the phase of evaluation not technical, but business. At this point, with the selected model, do we achieve the business objectives? The group that deals with this phase is a different group from the previous ones. It consists of resources that must *"use"* the system, feeding it with data different from those used in the previous phases.

If I have to question a student to assess whether he has *"understood"* what he has studied, I will ask him questions that have a different form from the text on which he has learned. The same for models: they must be tested with *"Test"* datasets that must not be derived from previous data. This condition, which is not trivial or sometimes easy to achieve, and must be foreseen in the very early initial phase. Not only do you have to use models



with new data, but here you also have to compare the technical result from the business result. In other words: the model responds well from the point of view of quality metrics, but is the final result, the one that reaches the “customer”, whether internal or external, what we wanted?

It is not uncommon for only at this moment to realize that something is wrong, to the point of discovering that we have to - see arrow in the figure - go back to the initial phase of “*Understanding the Business*”.

Such an eventuality should not be considered as an error, but as an iteration whose necessity becomes evident only at this moment. If, on the other hand, everything is as expected, we move on to the “*Implementation*” phase or - in more general terms - “*Operation*”

Implementation

Artificial intelligence can be developed on different platforms, which can be very dissimilar from each other. All companies have their own IT infrastructure and above all the related internal skills. Reconciling these two areas is of paramount importance. The choice of the operating platform therefore influences the development platform, and vice versa.

These are still software, but the choice of the platform necessary to put the developed Machine Learning models into operation may not be aligned with the usual technologies in the company.

This has such a significant impact that it must be dealt with at an early stage. In other words, I may have to choose some Machine Learning technologies, rather than others according to the internal platforms - and skills - of those who will then have to put the system into operation.

Once the requirements have been met, we move on to verifying the usability of the system by the “customer” - internal or external - end, user manuals and operating guides are prepared for the management of the infrastructure.

It must be evaluated whether to include the AI component in the ICT monitoring system, since an object, such as the AI model, which is not yet a “mature” part of the ICT infrastructure, enters the application chain.

We are on the air: how to proceed?

AI-based systems are by definition dynamic. Business objectives can vary, available data can change, new data can arrive that make us decide to exploit its information potential, or not. This perspective - not mandatory - is indicated by the circular arrow surrounding the diagram.

But this is a further novelty of AI projects which, being linked to changing data and conditions, can require this sort of continuous iteration over time.

Conclusion

A process like the one described is not just a procedure, but wants to offer a vision of the world. In this case it is a new world and therefore, inevitably, requires an effort, a departure from custom. The proposed procedure, the revised CRISP-DM, can support companies, especially SMEs, to start this new path, but one of great potential.

Keywords: Artificial Intelligence (AI); SME; CRISP-DM; Mining; Machine Learning; AI standard process; Experiential intelligence; Understanding of the Business; Understanding Data; Data Preparation; Modeling; Evaluation and Implementation; Business ecosystem; Process automation; Innovation in SMEs

(*) **Damiano Cosma De Angelis**

AI Engineer, Consultant & Trainer