

4. Progetto di ricerca e sviluppo

4.1 Titolo del progetto

“Smart Electronic Invoices Accounting” - SELINA

4.2 Sintesi della proposta

Il progetto “Smart Electronic invoices Accounting”, in breve SELINA, mira all’innovazione dei processi di contabilizzazione mediante l’impiego di tecnologie software che inglobano algoritmi di classificazione efficienti ottenuti come risultati della ricerca scientifica nell’ambito machine learning.

I processi e i sistemi software di supporto alla contabilizzazione hanno raggiunto, nel corso degli anni un elevato livello di maturità e, al contempo, hanno consentito ai produttori di tali sistemi di raggiungere un ottimo e consolidato livello di diffusione nel mercato. La sempre crescente pervasività delle nuove tecnologie ICT, dovuta soprattutto dalla diffusione dei dispositivi mobili e oggetti “smart” connessi alla rete, ha determinato, in tutti i domini applicativi, una forte esigenza di innovazione per i prodotti e servizi informatici che, a sua volta, ha contribuito all’incremento della competitività tra le imprese. Anche nello specifico dominio della produzione di sistemi e servizi di supporto alla contabilizzazione, la domanda di innovazione ha determinato la definizione di strategie efficaci per far evolvere in modo innovativo la propria offerta al fine di mantenere il portfolio di clienti e, potenzialmente, incrementarlo.

In tale contesto, uno degli aspetti maggiormente suscettibili di innovazione e la classificazione contabile delle voci di fatturazione, processo, ad oggi, non completamente automatizzato. Uno dei motivi principali che ostacolano la completa informatizzazione del processo è rappresentato dalla mancanza di uno standard di categorizzazione e associazione delle informazioni relative alle voci di fattura nelle corrispondenti voci contabili.

Spesso, ciascuno studio commercialistico adotta le proprie convenzioni modificandole e adeguandole al portfolio clienti, a specifiche esigenze del processo di gestione e trattamento di tali informazioni e ad esigenze contingenti. Il progetto SELINA si pone come obiettivo primario quello di realizzare un sistema intelligente di contabilizzazione automatica delle informazioni estratte da documenti contabili, siano esse dematerializzati o già in formato elettronico strutturato, in grado generare in modo automatico le corrispondenze con le voci contabili. Tale sistema sarà dotato di un componente tecnologico avanzato in grado di apprendere, attraverso opportune tecniche e modelli, il comportamento degli utilizzatori al fine di inglobare anche le loro preferenze e convenzioni. Il sistema sfrutterà lo stato dell’arte tecnologico e scientifico nel campo del machine learning e del data mining.

4.3 Descrizione degli obiettivi

Il progetto d’innovazione sintetizzato al punto precedente prende le mosse dall’evoluzione in corso nell’attività di commercialisti, studi professionali e consulenti del lavoro. L’ecosistema imprenditoriale italiano, composto per la maggior parte da micro e piccole imprese, senza il contributo di avvocati, commercialisti ed esperti contabili e consulenti del lavoro, non potrebbe svolgere direttamente molti importanti processi amministrativi, contabili e fiscali. I commercialisti si possono definire gli interlocutori per eccellenza delle aziende, indipendentemente dalla dimensione delle stesse, perché sono la prima figura professionale che si affianca a una PMI fin dalla sua nascita.

Le attività tipiche o tradizionali sono gestione della contabilità alle imprese (la effettua il 96% dei professionisti) e compilazione della dichiarazione dei redditi (svolta dall’84% degli studi italiani). Queste attività si caratterizzano per un elevato tasso di lavoro operativo (necessario a caricare i dati sul gestionale dell’impresa) e per un’elevata qualificazione dell’attività svolta (la scrittura contabile che viene effettuata presuppone conoscenze specifiche in campo contabile, amministrativo, fiscale).

Purtroppo sempre più spesso la percezione degli imprenditori nei confronti di queste attività “tradizionali” è quella di acquistare una “commodity”, ovvero di un servizio non differenziale per l’azienda e che può essere erogato da qualsiasi Commercialista in modo indifferente, o addirittura di mero adempimento burocratico senza nessuno scopo utile all’azienda.

La crisi economica ha reso necessario, tanto per le imprese quanto per i professionisti, recuperare livelli superiori di efficienza interna, riducendo gli sprechi e, soprattutto, ottimizzando i processi lavorativi, comprimendo le attività a basso valore. In tale ottica anche i commercialisti si sono visti costretti a ricercare nuove aree di business o nuovi canali di vendita con l’obiettivo di aumentare la fidelizzazione della clientela servita e aprire lo studio a nuove fasce di mercato attraverso nuovi servizi. Efficientare i processi interni significa per i professionisti avere più tempo da dedicare ai propri clienti e per sviluppare servizi nuovi.

Il progetto si pone in tale ottica cercando di innovare il processo gestione dei dati contabili attraverso l'introduzione di tecnologie che rappresentano lo stato dell'arte nel campo del machine learning e dell'analisi dei dati al fine di consentire una classificazione e gestione intelligente dei dati contabili sulla base delle informazioni storiche e delle peculiarità del settore di attività dell'impresa.

La ricerca industriale viene in aiuto per risolvere tale problema mettendo in campo tecniche innovative, algoritmi progettati e sviluppati appositamente all'interno del progetto per realizzare un modulo per la classificazione automatica delle informazioni presenti in documenti contabili utilizzando il cosiddetto approccio "Machine Learning". Esso si basa su un processo induttivo finalizzato a costruire automaticamente un classificatore per uno specifico dominio osservando le caratteristiche di un insieme di documenti che sono stati precedentemente classificati da esperti del dominio.

Visto che la classificazione di documenti si basa sulla semantica, e che la semantica di un documento ha una natura soggettiva si rischia che l'appartenenza di un documento a una categoria non possa essere decisa deterministicamente. Tale problema, conosciuto con il nome "Inter-indexer inconsistency", si verifica quotidianamente quando più esperti umani devono classificare un documento ma vi è un disaccordo. Obiettivo del progetto è proprio quello di definire un approccio innovativo in grado di codificare automaticamente la semantica delle informazioni in base ai criteri di valutazione di differenti esperti e fornire un motore che dinamicamente, in base all'analisi delle preferenze che l'esperto ha espresso in situazioni simili, sia in grado di classificare le informazioni dei documenti contabili affinché esse siano opportunamente organizzate per i processi successivi. La tecnica che si intende applicare è definita "Meta-learning" ed è finalizzata ad integrare differenti algoritmi di machine learning al fine di ottenere modelli predittivi i cui indici prestazionali superano quelli ottenuti per i singoli algoritmi di cui lo stesso modello si avvale. Tali tecniche contribuiranno a realizzare di un componente software ad elevate prestazioni che sarà in grado di ridurre i tempi e i costi dedicati ai processi di gestione documentale.

In particolare tra gli algoritmi e tecniche da integrare si prevede di utilizzare le tecniche di classificazione (supervisionata e non supervisionata) basate su Data Mining e sull'estrazione di regole associative pesate. Le regole associative pesate sono regole che sono significative non solo dal punto di vista della frequenza delle singole voci contabili (item), ma anche per il contenuto informativo di ogni item. Il contenuto informativo dei singoli item è un valore numerico non negativo detto peso, utilizzato per stabilire la significatività di ogni singolo item. In questo modo pesi elevati indicano termini molto informativi che andranno a mitigare la bassa frequenza di un item nel dataset, consentendo al sistema di estrarre regole rare (con bassa occorrenza) ma elevato contenuto informativo, fornendo nuova conoscenza non ottenibile con l'uso delle tecniche tradizionali. I vantaggi della estrazione di Regole Associate Tradizionali e Pesate da dati gestionali concerne essenzialmente: 1) l'estrazione di conoscenza in un formato fruibile dall'utente a partire dalla conoscenza che, sebbene presente nei dati, non è immediatamente disponibile all'utente; 2) il fatto che la conoscenza estratta è conoscenza azionabile a disposizione degli utenti anche non esperti del dominio del data mining, rispetto all'analisi manuale svolta correntemente per l'analisi dei documenti commerciali.

Obiettivo collaterale, ma non meno importante scientificamente e industrialmente è la realizzazione di un sistema scalabile in grado di sfruttare opportunamente le moderne architetture di calcolo parallelo/distribuito per rendere i processi di gestione documentale più efficienti, distribuire in maniera ottimale l'utilizzo delle risorse, garantire alte prestazioni anche con un numero significativo di documenti da organizzare o che presentino internamente strutture complesse.

4.4 Descrizione e qualificazione dei risultati attesi

Il progetto di ricerca industriale e sviluppo sperimentale, oggetto della presente proposta, permetterà al proponente di dotarsi di una tecnologia innovativa nel settore della contabilizzazione automatica delle informazioni estratte da fatture, siano esse dematerializzate o già in formato elettronico strutturato, capaci di generare in modo automatico le corrispondenze con le voci contabili, secondo l'apprendimento di opportuni modelli in grado di inglobare anche le preferenze (convenzioni) degli utenti finali.

In particolare, tale tecnologia includerà algoritmi avanzati di classificazione delle informazioni presenti nelle fatture e tecniche innovative di estrazione di regole associative tradizionali e pesate a partire dai dati di fatturazione.

Il sistema SELINA sviluppato nell'ambito della proposta presenterà funzionalità avanzate ed innovative capaci di automatizzare i processi di estrazione di conoscenza da documenti contabili perseguendo l'obiettivo di riduzione dei costi e dei tempi dei processi di gestione documentale.

Le innovazioni introdotte permetteranno di potenziare la capacità competitiva complessiva dell'azienda