



Ministero dello Sviluppo Economico

DOMANDA DI AGEVOLAZIONI FINANZIARIE
DECRETO MINISTERIALE 5 MARZO 2018 – CAPO III
PROCEDURA A SPORTELLO



SVILUPPO DI UN'INTELLIGENZA ARTIFICIALE PER
FAVORIRE LA SOSTENIBILITÀ DELLA SPESA CORRENTE
DELLE FAMIGLIE E L'AFFIDABILITÀ DEI CREDITI
(MACCHINA INTELLIGENTE RECUPERO CREDITI VER. 0)

PROPOSTA PROGETTUALE IDEATA E REDATTA DA

GTS Consulting Srl

21 GENNAIO 2019

Indice

Indice	i
I ELEMENTI DESCRITTIVI DEI SOGGETTI PROPONENTI	1
1 STRUTTURA ORGANIZZATIVA, PRODUTTIVA E DI RICERCA E SVILUPPO	3
1.1 Fire	3
Descrizione della struttura organizzativa	3
Indicazioni sul management aziendale	4
Descrizione della struttura produttiva e della struttura dedicata all'attività di R&S	5
Unità locale nella quale verrà realizzato il progetto	6
1.2 Expert System	6
Descrizione della struttura organizzativa	6
Indicazioni sul management aziendale	6
Descrizione della struttura produttiva e della struttura dedicata all'attività di R&S	7
Unità locale nella quale verrà realizzato il progetto	7
1.3 Università di Messina – Dipartimento di Economia	8
2 SETTORE DI ATTIVITÀ E CARATTERISTICHE DEL MERCATO DI RIFERIMENTO	9
2.1 Fire	9
Descrizione dell'attività svolta	9
Informazioni quantitative sul mercato di riferimento	9
Posizione del proponente nel mercato di riferimento	10
Previsioni di posizionamento nel triennio	11
Fatturato nell'ambito del settore produttivo oggetto della ricerca	11
Informazioni sul sistema competitivo e indicazione dei principali concorrenti	12
Principali accordi tecnici e/o commerciali, licenze e brevetti detenuti	12
2.2 Expert System	12
Descrizione dell'attività svolta	12
Informazioni quantitative sul mercato di riferimento	13
Posizione del proponente nel mercato di riferimento	14
Previsioni di posizionamento nel triennio	14
Fatturato nell'ambito del settore produttivo oggetto della ricerca	15
Informazioni sul sistema competitivo e indicazione dei principali concorrenti	15
Principali accordi tecnici e/o commerciali, licenze e brevetti detenuti	15
II ELEMENTI DESCRITTIVI DEL PROGETTO	17
1 TITOLO E DURATA DEL PROGETTO	19
2 AMBITO TECNOLOGICO	19

3	SINTESI	23
4	FINALITÀ	25
5	OBIETTIVO FINALE DEL PROGETTO	31
5.1	Caratteristiche e prestazioni	31
5.2	Principali problematiche tecnico-scientifiche e tecnologiche	32
5.3	Soluzioni tecnologiche previste	33
6	RESPONSABILE DEL PROGETTO	35
7	OBIETTIVI REALIZZATIVI DEL PROGETTO	37
7.1	OR1 – Machine learning per estrazione di conoscenza dai dati strutturati	39
	DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO	39
	ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE	41
	OR1.1 Identificazione dei dati, etichettatura dei dati e tipologie di analisi	41
	OR1.2 Algoritmi di ML non supervisionato	42
	OR1.3 Algoritmi di ML supervisionato	43
	TEMPI DI REALIZZAZIONE	43
7.2	OR2 – Profilo multidimensionale dell'interlocutore	43
	DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO	43
	ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE	45
	OR2.1 Creazione data base integrato	45
	OR2.2 Principali determinanti del default	46
	OR2.3 Principali Determinanti della LGD	46
	TEMPI DI REALIZZAZIONE	48
7.3	OR3 – Applicazione, testing e fine tuning degli algoritmi	49
	DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO	49
	ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE	50
	OR3.1 Definizione degli esempi per convalida e verifica sperimentale dei dimostratori	50
	OR3.2 Sviluppo e applicazione dei dimostratori	51
	TEMPI DI REALIZZAZIONE	51
7.4	OR4 – Creazione base dati con informazioni automaticamente estratte da contenuti non strutturati.	51
	DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO	51
	ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE	56
	OR4.1 Creazione Corpus file audio e documenti taggati	56
	OR4.2 Riconoscimento file audio (ASR)	57
	OR4.3 Analisi linguistica profonda dei testi	57
	TEMPI DI REALIZZAZIONE	59
7.5	OR5 – Estrazione dal testo di informazioni per la profilazione puntuale del soggetto analizzato	59
	DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO	59
	ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE	60
	OR5.1 Analisi del contesto	60
	OR5.2 Sviluppo del sistema	61
	OR5.3 Test sul sistema sviluppato e feedback	61
	OR5.4 Integrazione con gli altri moduli	62
	TEMPI DI REALIZZAZIONE	62
7.6	OR6 – Pilota del sistema NLP	62
	DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO	62

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE	64
OR6.1 Progettazione interfaccia grafica	64
OR6.2 Sviluppo ambiente dimostrativo e di test continuo dell'analisi linguistica	64
OR6.3 Test con diverse categorie di utenti	64
TEMPI DI REALIZZAZIONE	65
7.7 OR7 – Data base	65
DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO	65
ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE	67
OR7.1 Definizione delle basi dati da utilizzare	67
OR7.2 Pulizia dei dati	68
OR7.3 Sicurezza dei dati	68
OR7.4 Oscuramento dei dati	68
TEMPI DI REALIZZAZIONE	69
7.8 OR8 – Sistema utente	69
DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO	69
ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE	71
OR8.1 Metodologie di base	71
OR8.2 Analisi della conoscenza dei phone collector	71
OR8.3 Analisi della conoscenza degli home collector	72
OR8.4 Analisi della conoscenza dei legal collector	72
OR8.5 Analisi della conoscenza dei production manager	72
TEMPI DI REALIZZAZIONE	73
7.9 OR9 – Valutazione del sistema MIRC.0	73
DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO	73
ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE	74
OR9.1 Metodologie di base	74
OR9.2 Test semi-automatici	75
OR9.3 Test con simulazione	75
OR9.4 Analisi diretta	76
OR9.5 Interazione con agenti esterni	76
TEMPI DI REALIZZAZIONE	76
8 RISULTATO INTERMEDIO ATTESO DEL PROGETTO	77
9 RISULTATO FINALE ATTESO DEL PROGETTO	81
10 DIAGRAMMA TEMPORALE DEL PROGETTO	85
11 ULTERIORI INFORMAZIONI SULLE VOCI DI SPESA PREVISTE NEL PROGETTO	87
11.1 Fire	87
Attrezzature	87
Consulenze	87
Materiali e Forniture	88
11.2 Expert System	88
Attrezzature	88
Consulenze	88
Materiali e Forniture	88
11.3 Università di Messina	89
Consulenze	89
Materiali e Forniture	89

MIRC.0

11.4	Spese fuori area Obiettivo	89
12	ELEMENTI A SUPPORTO DELLA RICHIESTA DI MAGGIORAZIONE DEL CONTRIBUTO	91
III	ELEMENTI VALUTATIVI	93
1	CARATTERISTICHE DEL SOGGETTO PROPONENTE	95
1.1	CAPACITÀ TECNICO-ORGANIZZATIVA	95
	Fire Spa	95
	Expert System Spa	96
	Università di Messina – Dipartimento di Economia	99
1.2	QUALITÀ DELLE COLLABORAZIONI	101
	Descrizione del Dipartimento di Ingegneria dell'Università di Ferrara	102
	Descrizione GTS Consulting	105
	Collaborazioni internazionali dell'Università di Messina	105
2	QUALITÀ DELLA PROPOSTA PROGETTUALE	107
2.1	FATTIBILITÀ TECNICA DEL PROGETTO	107
	Adeguatezza delle risorse strumentali e organizzative	107
2.2	RISULTATI ATTESI	109
2.3	GRADO DI INNOVAZIONE	113
3	IMPATTO DEL PROGETTO	115
3.1	INTERESSE INDUSTRIALE	115
3.2	POTENZIALITÀ DI SVILUPPO	117
4	RISORSE FINANZIARIE PER LA REALIZZAZIONE DEL PROGETTO	119

I^a PARTE: ELEMENTI DESCRITTIVI DEI SOGGETTI PROPONENTI

1 STRUTTURA ORGANIZZATIVA, PRODUTTIVA E DI RICERCA E SVILUPPO

1.1. Fire

Descrizione della struttura organizzativa

Fire è nata nel 1992 come società recupero crediti Fire S.a.s. di Sergio Bommarito & C. Nei primi anni di attività, la società dava vita alla sua prima filiale a Catania, seguita a distanza di poco tempo da quella di Palermo. I successi conseguiti, in termini di percentuale di crediti effettivamente recuperati rispetto ai titoli ceduti all'agenzia, spingevano la dirigenza, nel 1999, a dare nuova veste giuridica a Fire. La S.a.s. veniva trasformata in società a responsabilità limitata con capitale sociale pari a £. 190.000.000. Nel 2000, in considerazione dell'aumento del fatturato, veniva sottoscritto nuovo capitale, passando da £. 190.000.000 ad € 260.000. La Fire assumeva l'attuale forma giuridica di società per azioni. Nasceva Fire S.p.A., società di riscossione e recupero crediti che, a partire dal 2001, si diffondeva capillarmente in tutta la penisola con le filiali di Vibo Valentia, Caserta, Bari, Roma e Milano. Nel 2006 è stato ampliato ulteriormente il ventaglio di servizi offerti attraverso la creazione dell'area Factor, specializzata nella gestione dei crediti provenienti da operazioni di factoring, e attraverso la stipula di proficue collaborazioni con le Pubbliche Amministrazioni per il recupero dei loro crediti. Nel 2006 il capitale sociale passava da € 260.000,00 a € 600.000,00. A fine 2006 si perfezionava un'operazione di fusione per incorporazione in Fire S.p.A. di SIF S.p.A., società di factoring, al fine di acquisire lo specifico known-how nel settore gestione crediti ed il portafoglio clienti della incorporata. In data 28 marzo 2017 è stata costituita la FB5 Investments S.r.l. che rappresenterà la nuova Holding dell'intero Gruppo Fire possedendo il 90% delle azioni della Fire Group S.p.A..

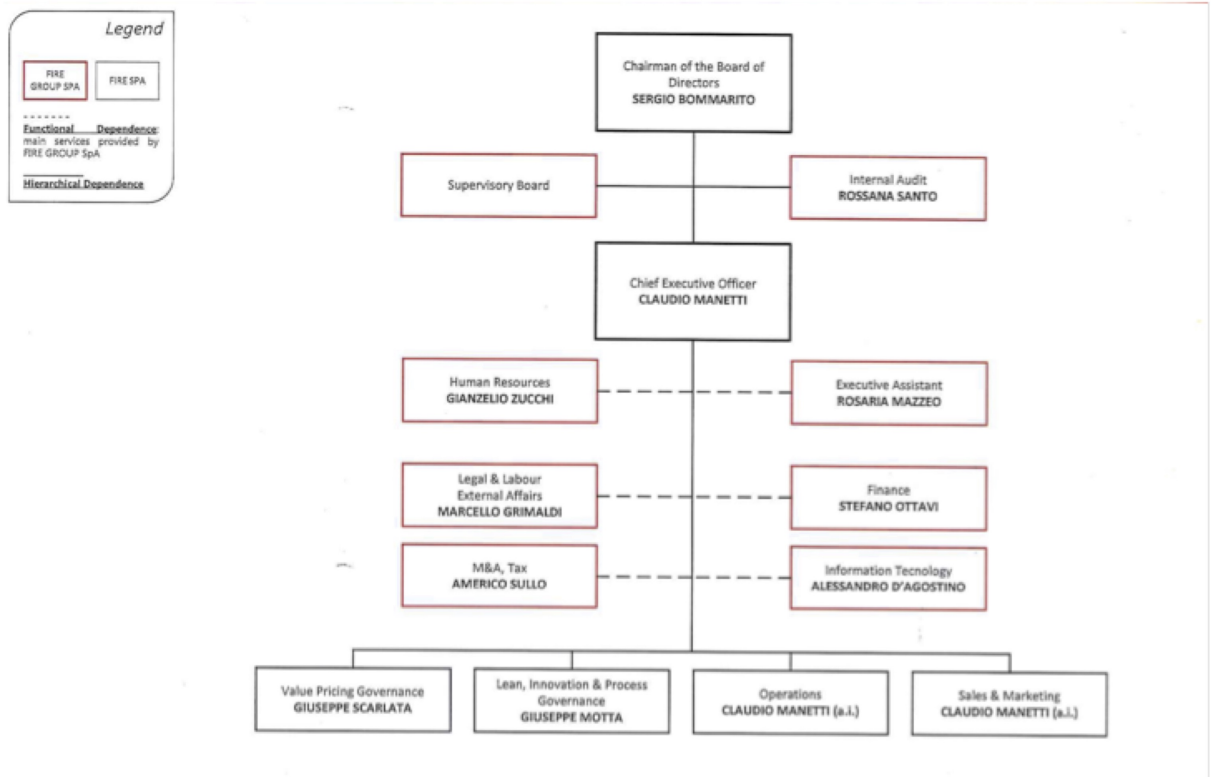
In data 6 aprile 2017 è stata costituita la Fire Resolution S.p.A., interamente posseduta dalla FB5 Investments S.r.l., avente ad oggetto, tra le altre, l'attività di consulenza, analisi e valutazione commerciale anche per conto terzi nelle operazioni di acquisto e di cessione di crediti garantiti o non garantiti e l'attività di gestione e recupero stragiudiziale di crediti ai sensi dell'art. 115 del TULPS.

In data 16 maggio 2017 si è perfezionato l'acquisto da parte della Fire Group S.p.A. del 34% del capitale della Fire Re S.r.l. A seguito delle predette operazioni straordinarie, la struttura del gruppo risulta come segue.



Indicazioni sul management aziendale

La Fire spa, ha come socio unico la Fire Group spa che la possiede interamente. E' amministrata da un consiglio di amministrazione presieduto da Sergio Bommarito, fondatore del gruppo e Claudio Manetti, Amministratore delegato.



Sergio Bommarito . MBA presso l'Istituto Superiore per Imprenditori e Dirigenti d'Azienda (I.S.I.D.A.) di Palermo (Business School for Entrepreneurs and Executives) e Laurea in Scienze bancarie e assicurative presso l'Università degli Studi di Messina, si citano i suoi principali ruoli: Membro fondatore di UNIREC (Unione Nazionale delle Imprese a tutela del credito); Membro fondatore ed ex membro del comitato esecutivo di A.I.I.R.E.C. (Associazione Italiana Imprese Recupero Crediti - Associazione italiana delle società di recupero crediti); Membro di A.C.M.I. (Associazione Credit Manager Italia - Associazione italiana gestori di credito); Direttore generale della Raccolta crediti certificato da CEPAS (ente di certificazione delle risorse umane - certificato da ACCREDIA - UNI CEI EN ISO / IEC 17024: 2012) registro n° 007 Membro del Comitato Esecutivo di Confindustria Servizi Innovativi, da maggio 2006 a maggio 2009. Claudio Manetti: dal 2015 ricopre il ruolo di AMMINISTRATORE DELEGATO di Fire, si occupa di Fatturato e sviluppo della redditività attraverso il rafforzamento del business esistente, cogliere opportunità come nuovi prodotti, mercati, M & A e perseguire la stabilizzazione e lo sviluppo di sistemi operativi. Laurea in scienze politiche ad indirizzo socioeconomico, ha ricoperto ruoli di rilievo nel settore marketing e sales di grandi compagnie quali: Fiatsava, Fiat group automobiles financial services, FGA Capital e Leasys Spa.

L'attuale organizzazione di Fire non è una struttura gerarchica per funzioni, ma un'organizzazione per linee di business, piccole aziende all'interno dell'azienda stessa, che beneficiano degli strumenti della grande ma sono più specializzate, flessibili, reattive e meglio coordinate. Tale organizzazione si basa su due figure chiave, il Manager della Linea di Business, responsabile del grado di soddisfazione del cliente e della redditività della propria linea di business, ed il Responsabile di Gestione del Mandato (R.G.M.) che risponde dei risultati del mandato conferitogli. Le linee di business rappresentano un core qualificato in grado di soddisfare le esigenze di qualsiasi cliente attraverso un processo di recupero specializzato sulle caratteristiche del prodotto e del debitore.

Descrizione della struttura produttiva e della struttura dedicata all'attività di R&S

La società Fira punta oggi a realizzare un processo di recupero crediti che sia un continuum operativo, grazie alla completa automazione delle attività di import, scoring e gestione ad eventi: l'anima di questo percorso sarà costituita dal cluster engine, un workflow realizzato dalla funzione di Information Technology che muove le posizioni da gestire tra gli operatori a seconda degli eventi che si manifestano, applicando regole personalizzate per ciascun cliente, per le diverse anzianità dei crediti, (le strategie). La Machine Learning è lo strumento che potrà essere utilizzato per dare uno score alle singole pratiche e clusterizzare le lavorazioni.

Allo scopo di poter implementare il nuovo processo, è stato eseguito un cambio di gestione molto profondo, che ha interessato l'organizzazione operativa. Ispirato al lean engineering, è stato codificato un nuovo metodo di lavoro, standardizzato nei processi, personalizzato nelle lavorazioni. Sono nate nella seconda parte del 2017 le Business Unit che rappresentano la ferma volontà di portare in azienda il mercato (ne esistono cinque, una per ogni comparto di mercato) e di rendere più omogenea ed efficace questa nuova modalità di lavoro. Ogni Business Unit ha al suo interno Value Stream, specializzate in relazione all'anzianità dei crediti da gestire, al cui interno operano i Product Manager, referenti dei nostri clienti. A questi ultimi spetta la responsabilità di risultato operativo, ai Value Stream Manager quella di risultato economico, ai Business Unit Manager quella di coordinamento e indirizzo, essendo in grado di rappresentare la sintesi dei clienti del settore gestiti su più fasi.

Trasversalmente alle Business Unit operano team specializzati: Legal, Analytics, Field Governance, Small & Medium Enterprises, Informazioni Commerciali e Back-office & Cash Management, al cui interno sono state collocate tutte le attività di carattere amministrativo inerenti la gestione dei mandati.

La struttura produttiva aziendale è composta al 30 novembre 2018, da 147 dipendenti dedicati alle seguenti aree:

- Operation
- Lean, Innovation and Process Governance
- Analytics
- Sales and marketing
- Value Pricing Governance

ai quali si affiancano:

- 807 collaboratori con il compito di effettuare attività di recupero crediti telefonico
- 203 collaboratori con il compito di effettuare attività di recupero crediti al domicilio
- 70 professionisti avvocati con il compito di effettuare attività di recupero crediti in fase giudiziale

La struttura dedicata alle attività di R&D si avvale della collaborazione dei reparti Process Governance, Innovation ed Analytics composte in totale da 12 figure professionali la cui formazione comprende le seguenti aree di competenza:

- Ingegneria Gestionale
- Statistica
- Matematica
- Informatica
- Economia
- Lean Organization
- Six Sigma
- Impact Innovation

Unità locale nella quale verrà realizzato il progetto

Il progetto sarà realizzato presso la sede di Messina, in via Bonsignore 1.

1.2. Expert System

Descrizione della struttura organizzativa

Expert System sviluppa software di cognitive computing e text analytics basati su algoritmi di intelligenza artificiale in grado di comprendere il significato del linguaggio scritto, in modo simile all'uomo. Aiuta le aziende di qualunque settore e le organizzazioni governative a gestire e sfruttare a pieno la quantità sempre maggiore di informazioni interne ed esterne, in particolare quelle contenute in testi non strutturati come pagine web, documenti, email, ecc.

La Expert System offre soluzioni per aumentare la conoscenza, prendere decisioni più informate, trasformare la customer experience, automatizzare i processi che fanno largo uso di informazioni e mitigare i rischi operativi. Le tecnologie tradizionali trattano il testo come una sequenza di caratteri, contano la frequenza delle parole o gestiscono elenchi di termini al fine di addestrare un sistema di corrispondenze, scaricando sul cliente l'onere della continua gestione del sistema. L'azienda, attraverso Cogito, tecnologia cognitiva, sfrutta sia l'intelligenza artificiale basata sulla semantica sia il machine learning per comprendere il testo simulando la comprensione umana. In questo modo, i sistemi e i processi che dipendono dall'elaborazione dei testi possono sfruttare l'intelligenza cognitiva per acquisire tutte le informazioni presenti nei testi, su vasta scala e senza sacrificare la precisione.

Dal 18 febbraio 2014 Expert System S.p.A. (ES) è quotata su AIM Italia, mercato di Borsa Italiana per le piccole e medie imprese italiane ad alto potenziale di crescita. Le risorse finanziarie raccolte con il processo di quotazione sono impiegate da Expert System per sostenere lo sviluppo e aumentare la visibilità dell'azienda, principalmente tramite tre linee di crescita:

- Espansione internazionale
- Sviluppo di partnership strategiche
- Consolidamento ed estensione di Cogito, la tecnologia proprietaria, brevettata negli USA.

Expert system è iscritta nella sezione speciale della CCIAA come PMI Innovativa.

Indicazioni sul management aziendale

Le principali figure del management aziendale sono:

Stefano Spaggiari Fondatore e CEO Da oltre 15 anni Stefano è Amministratore Delegato di Expert System, che grazie alla sua guida ha raggiunto una posizione di leadership e ha ottenuto prestigiosi riconoscimenti internazionali, affermandosi fra le migliori società ICT al mondo nell'ambito della gestione delle informazioni. Ha fondato l'azienda con l'idea, sempre portata avanti con grande passione e determinazione, di sviluppare un software per l'analisi delle informazioni a supporto delle aziende. Di recente Stefano ha gestito con successo la quotazione in Borsa di Expert System e guida il processo di internazionalizzazione dell'azienda. Stefano ha conseguito la laurea in Informatica all'Università di Modena e un EMBA presso l'ALMA Business School dell'Università di Bologna.

Marco Varone: fondatore, Presidente e CTO Marco, fondatore, Presidente e CTO di Expert System, è uno dei maggiori esperti mondiali di tecnologia semantica e di trattamento del linguaggio naturale. Ha ideato la piattaforma Cogito che è alla base di tutti i prodotti di Expert System e ha lavorato sul campo all'applicazione della tecnologia semantica in ogni contesto avanzato: motori di ricerca, text analytics, interfacce in linguaggio naturale, sistemi di Q&A, categorizzazione automatica e tanti altri ancora. Ha guidato l'azienda nella sua crescita costante e nel raggiungimento di importanti successi in diversi settori di business e coordina tutta l'attività di Ricerca e Sviluppo. Quando trova il tempo, scrive di questi temi sul proprio blog SemanticaMente e su quello aziendale.

Andrea Melegari SEVP, Defense, Intelligence & Security Andrea è Senior Executive Vice President, Defense, Intelligence & Security. In Expert System dal 2000, ha maturato grande esperienza nel realizzare e coordinare progetti molto complessi che impiegano la tecnologia semantica a supporto del processo di Intelligence. Dopo aver fondato e guidato per 12 anni l'Intelligence Division di Expert System, è stato uno dei promotori della joint venture con Elettronica che ha portato alla nascita di CY4Gate, nuovo player globale nel settore Cyber Electronic Warfare & Intelligence, di cui Andrea è Chief Marketing & Innovation Officer. Attualmente è membro del consiglio di Amministrazione di Expert System, di CY4Gate e di Expert System USA. Prima di far parte di Expert System, Andrea è stato consulente ICT, ha insegnato presso l'Accademia Militare di Modena per oltre 10 anni, e ha fondato la web agency Expertweb, acquisita in seguito dal Gruppo Banzai. Andrea ha conseguito la laurea in Informatica presso l'Università di Modena.

Descrizione della struttura produttiva e della struttura dedicata all'attività di R&S

Expert System S.p.A. ha sede legale a Rovereto (TN) e unità locali dislocate in diverse regioni:

- Rovereto (TN) – Sede legale, sviluppo R&D e servizi professionali
- Modena – Sede amministrativa, commerciale, sviluppo R&D e servizi professionali
- Roma – uffici commerciali, sviluppo R&D e servizi professionali
- Napoli - sviluppo R&D e servizi professionali
- Verona c/o Università degli studi di Verona - sviluppo R&D
- Milano – uffici commerciali

Attualmente l'azienda conta 143 dipendenti, di cui 112 laureati soprattutto laureati in ingegneria o materie linguistiche, suddivisi nelle seguenti funzioni aziendali:

SERVIZI PROFESSIONALI	71
R&D	35
AMMINISTRAZIONE - HR	15
IT	4
FACILITIES	2
SALES/MARKETING	16

Expert system si avvale inoltre della collaborazione di professionisti, società di consulenza specializzate e Università. Il continuo investimento in attività di ricerca e sviluppo è un elemento chiave della strategia di Expert System ed è necessario per restare allo stato dell'arte nel campo delle tecnologie cognitive applicate alle informazioni non strutturate ed essere così in grado di soddisfare le crescenti esigenze dei clienti e conservare il vantaggio competitivo nei confronti dei clienti. I costi sostenuti per tali attività sono stati capitalizzati, anche nel contesto di alcuni progetti di ricerca pluriennali nazionali ed internazionali cui l'azienda ha partecipato. Tali attività hanno riguardato la piattaforma semantica Cogito e i prodotti che sfruttano la piattaforma per indirizzare i più comuni casi di uso dei clienti. Per quanto riguarda gli strumenti di sviluppo utilizzati per la personalizzazione dei progetti, il maggior sforzo è stato indirizzato all'inserimento e alla messa a punto di tecniche di machine learning/deep learning, all'arricchimento delle funzionalità per la gestione del knowledge graph (Sensigrafo) e al completamento dell'integrazione con i componenti dello stack tecnologico rivenienti dall'acquisizione dell'attuale controllata francese.

Unità locale nella quale verrà realizzato il progetto

Il progetto sarà realizzato nelle unità locali di

- Modena – Sede amministrativa, commerciale, sviluppo R&D e servizi professionali;
- Napoli - sviluppo R&D e servizi professionali.

1.3. Università di Messina – Dipartimento di Economia

Al Dipartimento di Economia, costituitosi nel 2012, afferisce la quasi totalità del personale docente dell'Ateneo dei settori scientifico-disciplinari compresi nell'Area 13 "Scienze Economiche e Statistiche"; pertanto, esso si presenta come la struttura scientifico-didattica di riferimento dell'Università degli Studi di Messina per l'organizzazione di tutte le attività formative e di ricerca di tale Area. Il personale docente strutturato è nel complesso costituito da n. 67 unità (17 ordinari, 19 associati, 31 ricercatori). Da diversi anni il Dipartimento ha avviato attività scientifiche e formative con una forte proiezione di carattere internazionale grazie alle numerose collaborazioni che diversi docenti afferenti hanno con altre istituzioni accademiche e di ricerca sia italiane che straniere. Ciò al fine di creare presso l'Università degli Studi di Messina un polo di respiro internazionale nelle scienze economiche ed aziendali, attrattore di intelligenze (studenti e docenti) provenienti anche dall'estero.

L'obiettivo strategico del Dipartimento è quello di promuovere e potenziare principalmente la qualità della ricerca e di far discendere dall'eccellenza nella ricerca un'offerta didattica qualificata, di alto profilo e competitiva nel "mercato" della conoscenza. Il Dipartimento si propone anche di mettere a servizio del territorio le proprie competenze scientifiche che possono costituire un'importante risorsa, per le organizzazioni e le istituzioni locali, ai fini di una più rigorosa e approfondita conoscenza del territorio e del suo tessuto produttivo, così come al fine di delineare e progettare percorsi di sviluppo locale e di crescita dell'economia.

Presso il Dipartimento di Economia hanno sede:

- il "FLIRE. PHA (Fire Centre for Research on Payment Habits Analysis)", che persegue la finalità di promozione e svolgimento di attività di formazione e di ricerca, sia teorica che empirica, sui temi della finanza, con particolare attenzione alla gestione del risparmio e al recupero del credito;
- il Centro di Ricerca di Economia e Management Sanitario (CREMS), che ha come fine la promozione di programmi di alta formazione e di ricerca nel settore dell'economia e del management sanitario;
- il Pan.Lab., piattaforma scientifica dedicata al controllo della filiera produttiva agroalimentare limitatamente agli aspetti economico-finanziari.

2 SETTORE DI ATTIVITÀ E CARATTERISTICHE DEL MERCATO DI RIFERIMENTO

2.1. Fire

Descrizione dell'attività svolta

L'attività di gestione e recupero crediti stragiudiziale ha importanti ricadute sociali e, per essere svolta con professionalità, richiede regole certe, standard di qualità e rispetto reciproco a tutte le parti coinvolte: Società di Recupero Crediti, Azienda Committente, Cliente/Debitore. Fire S.p.A. è socio fondatore di UNIREC (Unione Nazionale Imprese di Recupero, Gestione e Informazione sul Credito, associata Confindustria Servizi Innovativi e Tecnologici e membro di FENCA – Federation of European National Collection Associations): adotta un codice deontologico interno che fissa regole precise di disciplina nei rapporti con i propri clienti e con i debitori. ha ottenuto, da parte di DNV GL, la Certificazione ISO 9001:2015 in relazione all'attività di esazione e recupero del credito ha ottenuto, da parte di DNV GL, la Certificazione ISO 27001:2013 in relazione alla gestione della sicurezza delle informazioni ha conseguito la Certificazione di oltre 1000 contratti a progetto da parte della Commissione istituita presso la Fondazione "Marco Biagi".

E' la prima società di recupero in Italia ad aver adottato un sistema di gestione per la conformità normativa "Compliance", secondo i criteri dettati dalla Banca D'Italia per gli intermediari finanziari. L'efficacia di tale sistema è certificata da DASA RÄGISTER.

Informazioni quantitative sul mercato di riferimento

Il sistema Paese ha chiuso il 2017 in netto miglioramento rispetto all'anno precedente e alle stime prudenti ancorché positive contenute nel Documento di Economia e Finanza redatto dal Ministero nel mese di Aprile 2017. In particolare quattro dinamiche hanno costituito un segno evidente del cambiamento di prospettiva della nostra economia:

- il Prodotto Interno Lordo è cresciuto del 1,5%, facendo registrare il miglior risultato dal 2010, grazie al trend sia della domanda interna sia dell'interscambio con l'estero; sulla prima hanno inciso favorevolmente gli incentivi fiscali indirizzati a stimolare gli investimenti in beni strumentali nonché l'aumento del valore aggiunto dell'industria, trainato dai settori manifatturiero e delle costruzioni;
- il rapporto deficit-Pil è sceso al 1,9%, il livello più basso da dieci anni a questa parte, pur non includendo questo dato la contabilizzazione degli effetti delle "Disposizioni urgenti per la liquidazione coatta amministrativa di Banca Popolare di Vicenza e di Veneto Banca";
- la pressione fiscale è diminuita di 0,3 punti percentuali, attestandosi al 42,4
- la disoccupazione ha raggiunto l'11,2%, mezzo punto in meno rispetto al 2016, consolidando il terzo calo consecutivo e il dato più basso dal 2013: quasi 300.000 in più sono stati i contratti a tempo determinato, 73.000 quelli a tempo indeterminato. Quest'ultimo fenomeno ha contribuito a innalzare il reddito disponibile, per altro ancora distante dal dato ante-crisi: coerentemente, sono cresciuti i consumi (specialmente di beni durevoli), mentre l'indebitamento delle famiglie si è ridotto al 61% del reddito disponibile così come la propensione al risparmio (pari al 8%). Sullo sfondo di queste dinamiche, il mercato del recupero dei crediti ha conosciuto un altro anno controverso. Sebbene non siano al momento ancora disponibili dati ufficiali, le ricerche condotte da società specializzate e

l'esperienza diretta maturata su un mercato sempre più competitivo possono contribuire a mettere in luce alcune fenomeni degni di nota emersi nel corso dell'anno:

- dal lato della domanda, è verosimile ritenere che l'ammontare complessivo dei crediti affidati al recupero si sia aggirato intorno agli 83 miliardi di euro, contro i 69,4 del 2016: una forte tensione lato volumi e una altrettanto robusta spinta verso l'aumento di complessità delle posizioni da gestire. Una crescita – seppure più modesta – dovrebbe aver riguardato anche il comparto finanziario del credito al consumo, attestandolo a circa 33 miliardi (+5,3%) per effetto delle lavorazioni prope-deutiche alle cessioni rolling dei crediti ante e post decadenza dal beneficio del termine. Qualche segnale di risveglio proviene invece dal leasing, settore poco orientato all'outsourcing del recupero dei crediti, il quale ha però visto letteralmente esplodere i volumi di new business grazie alla ripresa del settore automotive e alle agevolazioni finanziarie e fiscali riconosciute sugli investimenti in beni strumentali: poco meno di 700 milioni il valore dei crediti affidati all'esterno per il recupero, ma con attese di crescita. Una leggera diminuzione avrebbe invece caratterizzato il mondo variegato delle utility (energia, idrico e telecomunicazioni), ponendo fine al processo di assestamento seguito alle operazioni di riordino in particolare del comparto energetico: il 2017 si sarebbe così riposizionato intorno ai 7 miliardi di crediti affidati al recupero con una piccola flessione rispetto all'anno precedente. Tutto ciò che resta lato offerta (settori industriali, commerciali, assicurativo e pubblica amministrazione e privati) dovrebbero aver espresso un volume complessivo di affidamenti intorno ai 3,5 miliardi: un dato irrisorio laddove si pensi che solo i crediti delle aziende che non hanno incassato alla scadenza dalla pubblica amministrazione valgono circa 60 miliardi, che a sua volta il B2B ha portato a perdita nel 2017 crediti per ulteriori 60 miliardi e che affitti e spese condominiali non pagate si aggirerebbero intorno ai 14,5 miliardi di euro.
- Dal lato dell'offerta, il mercato ha osservato nel 2017 l'atteso processo di concentrazione e di consolidamento. Se da un lato gli investitori stranieri hanno perseguito e spesso attuato una strategia di integrazione verticale, acquisendo servicer nazionali, dall'altro si è manifestata la nuova tendenza a costruire piattaforme, tipicamente per operazioni di carve-out da parte degli istituti bancari

In sintesi, il mercato dei crediti da recuperare ha fatto registrare una crescita più che significativa, sebbene esistano segmenti ancora pressoché vergini (si pensi ai crediti commerciali): la concentrazione intorno agli NPL rischia di far perdere focus sul più tradizionale credito al consumo e sulle utenze. E' certo che in generale tutti i comparti offrono oggi crediti da recuperare molto più complessi, pur a fronte di commissioni sempre meno interessanti. C'è dunque una doppia sfida di fronte alla quale si trova l'intero comparto: far percepire il recupero dei crediti per ciò che è in realtà, ossia un regolatore sociale, una fonte di risparmio di risorse per le aziende che devono accettare di sostenere un costo superiore; la seconda sfida è intervenire sui propri mezzi di produzione, costruendo un nuovo paradigma operativo. In ossequio a queste necessità impellenti si dipanerà il prossimo triennio del settore.

Posizione del proponente nel mercato di riferimento

La Fire occupa il Quinto posto nella classifica de servicer italiani, ed è la prima società indipendente non quotata. La sua comprovata esperienza nella gestione di portafogli Unsecured e la presenza di Esperti Secured asset Manager, le hanno permesso di fare i seguenti numeri al 31/12/2017:

- 7,5 Mld in termini di AUM
- 6,9 Mld in termini di Due Diligence di 18 originator.

La tabella seguente mostra una classifica delle aziende di recupero crediti per fatturato, ed il posizionamento della Fire al quinto posto.

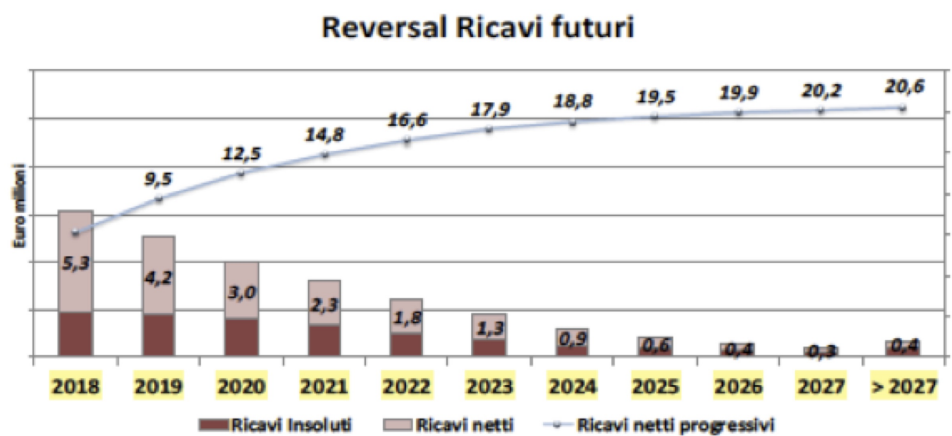
Table 6: Overview of main servicers (data at 31/12/2017) – Ranking by Revenues

Company	Bank of Italy surveillance	Revenues (€ million)	Special Servicing		Servicing Performing AuM (€ billion)
			Total Bad Loans ¹ AuM (€ billion)	Other NPL AuM ² (€ billion)	
doBank ³	Bank	213.0	74.3	1.5	0.9
Cerved Credit Management	106	94.8	34.9	1.4	9.5
MBCredit Solutions	106	70.3 ⁴	5.5	-	-
Credito Fondiario	Bank	41.6	6.3 ⁴	1.3 ⁴	1.4 ⁴
Fin	115	40.3	3.2	4.0	9.0
Advancing Trade	106/115	35.0	3.1	0.9	1.0
Guber	Bank	30.9	7.9	-	-
Crisis Credit Management	115	25.2	2.1	12.9	7.6
CAF (Intrum Italy)	115	22.7	8.2	-	0.2
Serfin	115	22.0	0.5	0.3	0.7
Hoist Italia	115	21.3	6.8	-	-
Securitisation Services	106	20.7	0.6	1.0	2.5
FBS	106	20.4	8.0	-	-
Sistemica	115	20.3	6.5	-	-
Europa Factor	106/115	20.1	2.6	-	0.5

Fonte: PwC, NPL Annual Report, Giugno 2018

Previsioni di posizionamento nel triennio

Di seguito è rappresentato il piano di reversal dei predetti “Ricavi futuri” in funzione delle scadenze dei titoli/effetti a supporto dei piani di rientro sottoscritti



Per il prossimo triennio la Società ritiene di mantenere lo stesso trend degli ultimi anni.

Fatturato nell’ambito del settore produttivo oggetto della ricerca

L’esponentiale crescita del ramo Bad Loans ed Unlikely to pay, e più in generale dei Non Performing Loans, che ha caratterizzato il business aziendale negli ultimi anni, ha generato un’importante crescita degli incassi effettuati tramite Piani di rientro (di seguito “Pdr”) andando, pertanto, ad alimentare una «riserva» di ricavi per Pdr il cui reversal si manifesterà nei prossimi esercizi.



Di seguito i numeri del 2017:

<i>Euro migliaia</i>	2017	2016	Variazione	Variazione %
N. pratiche gestite (in migliaia)	4.514	4.042	472	11,7%
GBV crediti affidati (in milioni)	€ 16.193 mln	€ 15.148 mln	€ 1.045 mln	6,9%
Ricavi	€ 40.316 k	€ 40.322 k	(€ 6 k)	(0,0%)
EBITDA	€ 1.145 k	€ 1.397 k	(€ 252 k)	(18,0%)
EBIT	€ 948 k	€ 1.220 k	(€ 272 k)	(22,3%)
Risultato netto	€ 478 k	€ 605 k	(€ 127 k)	(20,9%)
Flusso di cassa alla fine dell'esercizio	€ 4.074 k	€ 3.211 k	€ 862 k	26,5%
Posizione finanziaria netta	€ 1.760 k	€ 2.004 k	(€ 244 k)	(12,2%)

Informazioni sul sistema competitivo e indicazione dei principali concorrenti

Principali accordi tecnici e/o commerciali, licenze e brevetti detenuti

Fire offre i propri servizi e le proprie soluzioni di Gestione e Recupero Crediti ad aziende leader nei seguenti settori: Finanziario, Bancario, Utilities, Telecomunicazioni, Commerciale, Cessione del Quinto, ed alle Pubbliche Amministrazioni. Fire interviene in ogni segmento di mercato soluzioni specifiche team di lavoro specializzati e di grande esperienza. Tutti i progetti, creati insieme alle committenti, sono finalizzati alla gestione ottimale dell'intero ciclo di recupero del credito e customizzati in funzione delle caratteristiche peculiari dell'area di business di riferimento. Di seguito alcuni dei principali clienti:

- Telecom Italia
- Sorgenia
- Enel Energia S.p.A.
- Compass SpA (Gruppo Bancario Mediobanca)
- Fidelity

2.2. Expert System

Descrizione dell'attività svolta

Expert System è stata fondata a Modena da tre colleghi universitari con l'obiettivo di dimostrare che anche in Italia è possibile creare software di alto livello. All'epoca, la convergenza tra linguistica e tecnologia era qualcosa di cui si parlava solo negli istituti di ricerca o negli ambienti accademici. Dopo il primo contratto con Microsoft, l'azienda ha avuto l'opportunità di dedicare tutte le risorse allo sviluppo

di un software di intelligenza artificiale in grado di capire il significato del linguaggio in base al contesto. Lo sforzo ha prodotto una delle prime piattaforme di analisi semantica: la tecnologia brevettata Cogito.

Nel 2014 La Expert System ha fatto il suo ingresso nel mercato azionario italiano e subito dopo ha completato l'acquisizione di due divisioni dell'azienda spagnola di tecnologia semantica ISOCO e del leader della text analytics TEMIS. Ha lanciato CY4GATE, azienda che opera nel settore dell'elettronica per la cyber security internazionale, e rafforzato la sua presenza negli Stati Uniti tramite una nuova struttura organizzativa per far fronte alle crescenti necessità dei clienti del settore pubblico e privato.

Informazioni quantitative sul mercato di riferimento

L'Artificial Intelligence sta rivoluzionando il modo in cui lavorano le aziende, alle quali offre una concreta modalità di successo: essa infatti consente di andare oltre una gestione puramente "numerica" dei dati, aggiungendo la capacità di comprendere e gestire le informazioni per ricavarne il massimo valore di business.

Nel mondo moderno, è di vitale importanza per le organizzazioni di qualsiasi dimensione disporre della capacità di creare, catturare, archiviare, individuare e condividere la conoscenza, ma anche disporre di adeguate piattaforme che consentano di usufruire delle informazioni necessarie al momento giusto. La tecnologia semantica, con la sua capacità analoga a quella della mente umana di capire il significato delle parole in base al contesto, supporta individui ed organizzazioni nel cruciale processo di gestione delle informazioni disponibili, e sempre più spesso consente alle imprese di rispondere rapidamente ai cambiamenti del business, ottenere informazioni approfondite su clienti, industrie e mercati, e diventare soggetti più proattivi.

La maggiore consapevolezza della opportunità offerte dai Big Data si riflette a sua volta nella pianificazione di strategie aziendali data driven, orientate agli aspetti predittivi ed all'automatizzazione di processi e servizi, ed in conseguenti investimenti dai ritorni sempre più certi e misurabili, avvalendosi ove necessario di nuove e specializzate figure professionali quali i data scientist. Il settore della Big Data Text Analytics in cui opera Expert System continua pertanto a mantenere considerevoli prospettive di crescita, sia a livello globale dove nel corso del 2017 si è attestato a quota 150,8 miliardi di dollari, che nel mercato interno dove è stata superata la soglia del miliardo di Euro.

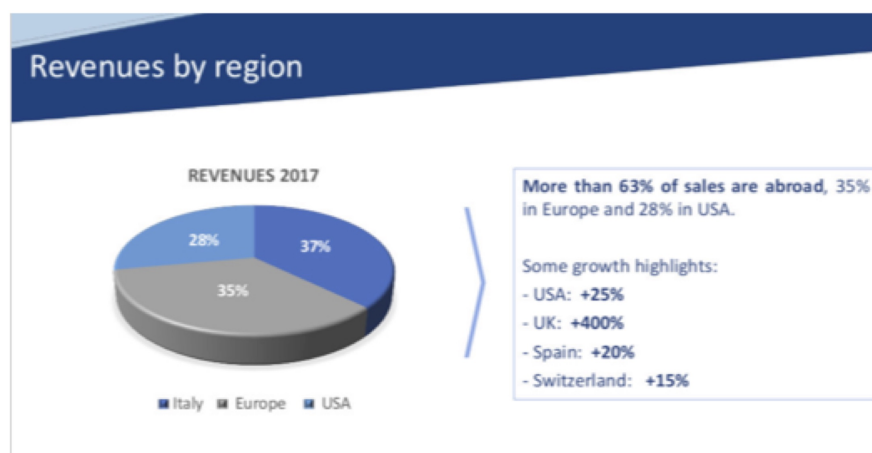
Tale crescita ha interessato i più svariati settori di attività, ovvero quello bancario e assicurativo, manifatturiero, telco e media, PA e sanità, servizi, Gdo e utility, a testimonianza di come l'opportunità di cogliere risultati promettenti sia trasversale rispetto all'attività economica di riferimento. Expert System ha toccato con mano tale notevole interesse in occasione del primo Cogito AI Day, da essa organizzato a Milano in data 19/10/2017, nell'intento di condividere con aziende, analisti ed esperti di settore i reali vantaggi e gli utilizzi concreti dell'intelligenza artificiale, nonché l'evoluzione del mercato di queste tecnologie. Nel 2017, il mercato digitale italiano è cresciuto del 2,3% 68.722 milioni di euro e lascia intravedere ancora crescita per i prossimi anni: 2,6% per il 2018, 2,8% per il 2019, 3,1% per il 2020. Il trend discendente degli anni più bui della crisi appare oramai alle spalle sull'onda delle componenti più legate all'innovazione. Il gap digitale accumulato in passato obbliga a un passo ancora più sostenuto, centrato sull'accelerazione delle politiche per l'innovazione già avviate, per l'ammodernamento della pubblica amministrazione, l'inclusione digitale delle piccole imprese e dei territori, lo sviluppo diffuso delle competenze. Queste le principali evidenze delle rilevazioni di Anitec-Assinform - l'Associazione di Confindustria che raggruppa le principali aziende del settore - condotte in collaborazione con NetConsulting cube.

Guardando i dati più in dettaglio, alla crescita del 2017 hanno concorso tutti i macrocomparti, fatta eccezione per le infrastrutture immateriali, che però hanno tenuto a 22.346 milioni (-0,1%), arrestando un calo che durava da anni. I servizi ICT hanno raggiunto 11.056,8 milioni (+4%), il software e le soluzioni ICT 6.626,1 milioni (+5,9%), i dispositivi e sistemi 18.332,7 milioni (+0,2%), i contenuti digitali e digital advertising 10.360,3 milioni (+7,7%). Ovunque gli andamenti dei diversi segmenti hanno una volta di più confermato il rallentamento delle componenti più tradizionali e il dinamismo di quelli più legati alla trasformazione digitale dei modelli produttivi e di servizio. I macrocomparti del Software e Soluzioni ICT e dei servizi ICT sono quelli che più hanno animato la trasformazione in atto, beneficiandone.

Posizione del proponente nel mercato di riferimento

Attraverso la tecnologia proprietaria Cogito, basata sull'analisi semantica, la Expert System innova il modo con cui le aziende elaborano le informazioni, offrendo il proprio tool di intelligenza artificiale per comprendere il significato di grandi quantità di documenti, e ricavare conoscenza strategica dai big data per accrescere il proprio vantaggio competitivo e migliorare i processi decisionali. Oggi, l'azienda è leader nel campo del cognitive computing e della text analytics e offre un numero sempre crescente di soluzioni in diverse lingue a molte delle più importanti aziende e agenzie governative in Europa, Stati Uniti e Medio Oriente. Fra i principali clienti del gruppo, che ha sedi in Italia, Francia, Spagna, Germania, Gran Bretagna, Stati Uniti e Canada figurano: Shell, Chevron, Gruppo Eni, Telecom Italia, Intesa Sanpaolo, Il Sole 24 Ore, Ely Lilly, BAnQ, Biogen, Bloomberg Bna, Elsevier, Gannett, Thomson Reuters e Wolters Kluwer, oltre al dipartimento dell'Agricoltura e a quello della Giustizia degli Stati Uniti.

Il punto di forza della società è la capacità di investire: 6 milioni, ovvero oltre un quinto del fatturato, ogni anno. In questo modo Expert System è passata in quattro anni dagli 8 milioni di fatturato per l'85% realizzato in Italia agli oltre 26 milioni, per due terzi derivanti dai mercati esteri. Dal 2016 al 2017 L'expert System è passata da un ebitda negativo per due milioni a un ebitda positivo per un milione. Tre milioni di miglioramento sui margini, tre milioni che corrispondono all'aumento di fatturato nel 2017, a parità di costi, in quanto si cominciano a vedere i frutti della vendita di licenze. Il continuo investimento in attività di ricerca e sviluppo è un elemento chiave della strategia di Expert System ed è necessario per restare allo stato dell'arte nel campo delle tecnologie cognitive applicate alle informazioni non strutturate ed essere così in grado di soddisfare le crescenti esigenze dei clienti e conservare il vantaggio competitivo nei confronti dei concorrenti.



Previsioni di posizionamento nel triennio

L'azienda si pone obiettivi ambiziosi per i prossimi anni. Obiettivi che vanno da una maggiore penetrazione nei mercati mondiali in cui è già presente, in particolare Usa, Francia e Germania; al rafforzamento della propria posizione di vendor anche attraverso partner e system integrator, per concentrarsi sul miglioramento di margini e profittabilità dei progetti, arrivando all'utile netto nel 2019 (fino ad oggi, la scelta è stata di investire il più possibile in ricerca). Senza escludere nuove acquisizioni, dopo quelle che, negli ultimi 4 anni, dopo la quotazione su Aim, l'hanno portata a internazionalizzarsi. Expert System S.p.A. continuerà ad essere l'azienda guida sul fronte ricerca e sviluppo all'interno di un gruppo che va via via sempre più internazionalizzandosi. Continua l'opera di sviluppo della tecnologia e soprattutto dei prodotti verticalizzati, che consentiranno all'azienda di offrire soluzioni specifiche e puntuali alle esigenze dei clienti in modalità sempre più standardizzata con chiari ritorni anche sulla redditività dei progetti.

I risultati annuali 2017, sia a livello di gruppo, sia livello di Expert System S.p.A., dopo un primo semestre debole, sono stati molto buoni e questo pone le basi per un consolidamento degli stessi per gli anni a venire. Il mercato italiano, pur non avendo le potenzialità e le dimensioni di alcuni mercati esteri in cui operano le filiali dell'azienda, mostra comunque una buona tenuta e interessanti opportunità di

sviluppo, sia nei settori più maturi (finanziario, assicurativo, media e governativo), ma anche su quello manifatturiero che caratterizza il mondo produttivo italiano.

Fatturato nell'ambito del settore produttivo oggetto della ricerca

Nella tabella che segue sono indicati i risultati conseguiti negli ultimi due esercizi in termini di valore della produzione, margine operativo lordo e risultato prima delle imposte.

	31/12/2017	31/12/2016
Valore della produzione	19.254.068	16.790.322
Margine operativo lordo	3.309.453	1.320.094
Risultato prima delle imposte	(1.277.476)	(1.474.922)

Il valore della produzione, pari a 19,3 milioni di Euro, cresce di oltre il 14% rispetto all'esercizio precedente, chiuso con un ammontare pari ad Euro 16,8 milioni. Tale incremento, unitamente alla ridotta crescita dei costi esterni operativi (+ 1,5%) e dei costi del personale (+4,7%) produce un consistente miglioramento della marginalità rispetto al precedente esercizio, con un EBITDA di Euro 3,3 milioni in crescita di oltre il 150% rispetto al dato di Euro 1,3 milioni registrato nel 2016. Il miglioramento del risultato ante imposte, negativo per Euro 1,3 milioni (negativo per Euro 1,5 milioni al 31/12/2016) è meno significativo dal momento che sconta una crescita degli ammortamenti (+25%), strettamente connessa ai maggiori investimenti in attività di sviluppo, ed è gravato da componenti non monetarie quali la valorizzazione di poste di bilancio in valuta diversa dall'Euro, sostanzialmente legate a contratti di finanziamento intercompany, che incidono negativamente per Euro 1 milione.

Informazioni sul sistema competitivo e indicazione dei principali concorrenti

Principali accordi tecnici e/o commerciali, licenze e brevetti detenuti

L'azienda è proprietaria di un brevetto: numero domanda US 11/744, 479 – data pubblicazione 1/3/2011 – data registrazione 4/5/2007 – data priorità 4/5/2007). Il brevetto, denominato “METHOD AND SYSTEM FOR AUTOMATICALLY EXTRACTING RELATION“ BETWEEN CONCEPT“ INCLUDED IN TEXT”, è il pluripremiato e multilingue Cogito®, il software per la text analytics e il cognitive computing che legge, capisce e risponde.

Cogito supporta le aziende a prendere decisioni migliori alla velocità delle tecnologie di oggi con la scalabilità necessaria per gestire grandi archivi di dati. Nell'ultimo anno l'attività di Expert System sul fronte tecnologico ha consentito di offrire al mercato una nuova release del motore di tecnologia semantica Cogito, in grado di semplificare ulteriormente lo sviluppo di soluzioni avanzate per l'automazione dei processi aziendali (roboticprocess automation) e l'information intelligence. Cogito 14, l'ultima versione rilasciata dai Cogito Labs, offre oggi ai clienti la miglior tecnologia di analisi semantica attraverso i prodotti di punta Cogito Studio e Cogito Discover. Da un lato, Cogito Studio Platform si propone come un sistema integrato per progetti di analisi testuale, ed include Cogito Studio, un editor di applicazione semantica che permette di valutare l'efficacia della categorizzazione o configurare l'estrazione da raccolte di documenti, misurando precisione e recall, e Cogito Studio Express, ovvero un editor di ontologie che fornisce una modalità semplice ed accessibile per importare thesaurus e vocabolari del cliente, prevedendo come l'ontologia si applichi al contenuto. Dall'altro, Cogito Discover è invece una piattaforma scalabile per la generazione automatica di metadati semantici e autoclassificazioni che può essere integrata in qualsiasi applicazione e flusso di lavoro implementabile su una piattaforma cloud o su ambienti ibridi. Cogito si conferma essere la tecnologia di analisi testuale alla base di tutti i prodotti forniti da Expert System, questa volta arricchito nel numero di lingue supportate, che comprende ora il portoghese, il russo e l'olandese.

Cogito 14 supporta oggi 14 lingue ovvero italiano, inglese, spagnolo, francese, tedesco, portoghese, russo, coreano, cinese, giapponese, arabo, greco, polacco e olandese.

II^a PARTE: ELEMENTI DESCRITTIVI DEL PROGETTO

1 TITOLO E DURATA DEL PROGETTO

Titolo del progetto MIRC.0 – Sviluppo di un’intelligenza artificiale per favorire la sostenibilità della spesa corrente delle famiglie e l’affidabilità dei crediti (Macchina Intelligente Recupero Crediti ver. 0)

Durata del progetto 36 mesi

2 AMBITO TECNOLOGICO

KET fondamentale: Tecnologie dell’informazione e della comunicazione (TIC)

Settore applicativo: Fabbrica intelligente

Tematica e sotto-tematiche

- SN A1 Processi produttivi ad alta efficienza e per la sostenibilità industriale
 - Soluzioni ICT per la valorizzazione e condivisione della conoscenza all’interno delle fabbriche
- SN A2 Sistemi produttivi evolutivi e adattativi per la produzione personalizzata
 - Strumenti avanzati per la configurazione e progettazione di soluzioni personalizzate
 - Strumenti per il supporto alle decisioni in ambienti complessi

Descrizione degli elementi del progetto con la tecnologia indicata La quarta rivoluzione industriale, nella visione di Industry 4.0, è guidata dall’introduzione nella fabbrica di tecnologie ICT avanzate, tra cui quelle basate su Intelligenza Artificiale (IA).

Il progetto si colloca alla frontiera degli impieghi di tecnologie di IA in campo industriale. Ha l’obiettivo di calare tecnologie altamente strategiche (in particolare Sistemi a Regole, Natural Language Processing, Machine Learning e analisi automatica dei dati) nel contesto del recupero del credito. Si tratta di ottenere, attraverso queste tecnologie, una serie di vantaggi competitivi nel campo della tutela del credito, che possono a medio termine produrre modifiche strutturali al modello di business.

Le tecnologie di IA cui si intende ricorrere sono molteplici: il sistema di supporto alle decisioni si configura come un sistema basato su conoscenza (knowledge-based), dove la conoscenza di dominio da formalizzare è ottenuta sia dagli esperti e dagli operatori di Fire, sia estraibile dalle conversazioni raccolte in registrazioni, e dai dati archiviati relativi agli interlocutori e alle pratiche pregresse.

Le tecnologie abilitanti identificate come utili al progetto sono pertanto tre:

- la formalizzazione e l’applicazione di regole, in un sistema knowledge-based (KB);
- le tecnologie di Natural Language Processing (NLP);
- le tecnologie e i metodi di Machine Learning (ML).

Il logo ideato per il sistema MIRC.0 vuole evidenziare l'orchestrazione di queste tecnologie. Il Decision support system, basato su regole, è rappresentato da un poligono in background: alla formalizzazione delle sue regole concorrono sia gli esperti di Fire (i vari lati del poligono), sia le tecnologie di NLP e ML, rappresentate dai due triangoli in esso innestati.



La conoscenza formalizzata in MIRC.0 si compone quindi sia del know-how e delle competenze umane (regole fornite da personale Fire) sia della conoscenza estratta dai dati strutturati e non raccolti (i due triangoli innestati nel logo), in particolare estratta con tecniche di NLP dalle conversazioni e appresa con tecniche di ML dai dati disponibili su interlocutori e pratiche.

Si descrivono le tre tecnologie più in dettaglio.

Sistemi a regole La prima tecnologia abilitante è quella dei sistemi a regole, ove le regole sono formalizzate in un opportuno linguaggio, e applicate da un motore inferenziale. In particolare, un sistema di supporto alle decisioni (DSS) basato su regole è un sistema decisionale che aiuta anche i principianti a prendere decisioni come un esperto. L'idea che sta alla base degli strumenti decisionali basati su regole è quella di mettere le conoscenze disponibili in letteratura e le competenze che appartengono ad alcuni esperti e renderle disponibili agli altri. Quindi sono spesso anche chiamati sistemi esperti o sistemi basati sulla conoscenza.

La base di conoscenza (knowledge base, o KB) di un DSS è composta solitamente da regole, dette anche regole di produzione, costituite da un antecedente e un conseguente, del tipo: if (antecedente) then (conseguente). L'antecedente può essere una somma di condizioni da verificare sui dati a disposizione, e il conseguente una somma di azioni da effettuare (ad esempio, se si tratta del primo contatto operatore-interlocutore, verificare la consapevolezza dell'interlocutore e non proporre alcuna ulteriore rateizzazione del credito).

All'aggiunta di un nuovo caso e nuovi dati di tipo fattuale rappresentati nel sistema (in quella che solitamente è chiamata memoria di lavoro), le regole sono applicate in avanti (dall'antecedente al conseguente) in modalità forward-chaining, da un motore inferenziale (che potrebbe – nei casi più semplici, deterministici – essere anche un motore di workflow).

La conoscenza formalizzata in regole viene quindi applicata al presentarsi di nuove informazioni, in una particolare situazione, in modo da consentire agli operatori di fare scelte più efficaci. I vantaggi di questo approccio sono che

- la conoscenza viene esplicitata e formalizzata in regole, in un linguaggio formale, simbolico e dichiarativo;
- le regole formalizzate sono validabili dagli esperti;
- il modulo di controllo (motore inferenziale) è separato e già disponibile;
- la manutenzione della base di conoscenza del sistema e il suo aggiornamento risultano facilitati, in quanto si interviene solo sulle singole regole.

Il collo di bottiglia di questo approccio knowledge-based è costituito spesso dalla capacità degli esperti di spiegare la loro metodica di lavoro, e la conseguente difficoltà di rappresentarla in regole. Poiché è noto che gli esperti prendono decisioni migliori rispetto ai principianti, e una delle ragioni è che hanno una migliore capacità di identificare, classificare e riconoscere un nuovo caso, riconducendolo a casi già visti e/o alla metodica di lavoro appresa, da letteratura o con l'esperienza, nel progetto si applicheranno due

ulteriori tecnologie abilitanti, al fine di apprendere, dalle conversazioni pregresse, dai dati di interlocutori e loro pratiche, ulteriori “regole” da integrare nella KB del sistema.

Il Machine Learning (ML) e il Natural Language Processing (NLP) sono due degli hot topic in Intelligenza Artificiale. Tutte le più importanti compagnie di ICT al mondo stanno investendo denaro e stanno assumendo persone con un forte background in queste tecnologie, perché oggi – specie dai dispositivi connessi a Internet – si raccolgono moltissimi testi e dati e si ha bisogno di ricavare valore da essi. Questo è vero anche nel caso di Fire, il patrimonio di informazioni raccolte (conversazioni registrate, e dati strutturati e non) è una notevole fonte informativa, da cui estrarre conoscenza con tecniche automatiche.

Natural Language Processing Il Natural Language Processing (NLP) è il trattamento automatico di informazioni testuali scritte in linguaggio naturale, di tutti i giorni, che a sua volta si compone di più fasi:

- Procedura di Entity Extraction (NER): si tratta del processo mediante il quale vengono automaticamente riconosciute entità presenti nel testo come: persone, luoghi, organizzazioni, procedure, ecc.. Tali informazioni vengono rese disponibili ad applicazioni di terze parti in forma strutturata (tipicamente in formato XML), abilitando quindi nuove ed esclusive analisi e viste sui dati.
- Procedura di Categorizzazione automatica: il cosiddetto “robot” ⁽¹⁾, è in grado, durante l’analisi del testo, di classificare (coding) i testi analizzati secondo specifiche tassonomie di dominio al fine di contestualizzare il contenuto e permetterne una fruizione attraverso filtri di ricerca.
- Writeprint: come l’impronta di un polpastrello (fingerprint), anche un testo riporta alcune nostre caratteristiche esclusive, attributi del linguaggio utilizzato che possono essere utilizzate a supporto della clusterizzazione (ad esempio uso di aggettivi, forme verbali complesse, ecc.) o profilazione degli utenti.
- Mood: è il tono della conversazione, calcolato non solo basandosi sulla presenza/assenza di certe keyword (esempio lista di termini scurrili), ma sull’utilizzo di particolari parole/concetti espresse anche in forme lessicali complesse (es. “penso proprio di essere preso in giro”, “ma mi stai ascoltando o no!”,...).

Machine Learning Il Machine Learning (ML) è oggi una delle principali tecnologie di IA, utilizzato estensivamente ad esempio nell’ambito della profilazione dei clienti di e-commerce. I dati disponibili possono essere valorizzati, con il ML e l’analisi intelligente, estraendo da essi conoscenza e profitto, applicando queste tecniche (per citare alcuni esempi applicativi della Internet economy, queste tecniche sono usate per migliorare la user-experience, per il retail, per la proposta di acquisti, etc).

Più in dettaglio, il ML è l’insieme di tecniche automatiche di analisi dei dati con cui si apprendono “modelli”, a fini classificatori (quindi per classificare nuovi casi non visti), per fare previsioni, o per raggruppare dati e individuarne caratteristiche salienti.

Ci sono due grandi approcci nel ML: il primo detto ML supervisionato, che, a partire da esempi, è in grado di apprendere un classificatore in modo automatico (ad esempio, il sistema di classificazione delle mail, in spam o non spam, adottato nel sistema di posta Gmail di Google), o di fare previsioni future sulla base di dati di input-output pregressi come nel caso della regressione.

Nel ML supervisionato, i dati disponibili sono etichettati con valori per una o più classi. I dati etichettati sono divisi in due parti, il training set e il test set con cui, rispettivamente, si addestra e si misura la performance del classificatore appreso. Tra le tecniche di ML supervisionato, si considereranno quelle che apprendono alberi di decisione, e regole di produzione (queste facilmente integrabili nella KB del sistema a regole).

Il secondo approccio, invece, detto ML non supervisionato, senza avere esempi pregressi a disposizione, è in grado di raggruppare i dati in ingresso, al fine di trovare le caratteristiche comuni, più salienti, ridurne la dimensionalità, etc.

¹Un robot (spesso abbreviato in “bot”) in ambito informatico è in senso lato un programma che accede alla rete attraverso lo stesso tipo di canali utilizzati dagli utenti umani per svolgere in maniera automatica compiti che sarebbero troppo gravosi o complessi per gli utenti umani.

Tra le tecniche di ML non supervisionato, il clustering ha l'obiettivo di selezionare e raggruppare elementi omogenei, in base a una misura di distanza o similarità. Diventa quindi fondamentale individuare le variabili (attributi) maggiormente significative per la mole dati a disposizione, sia definire la misura della distanza, in modo da tenere conto delle caratteristiche dei dati. Si considereranno tecniche di clustering quali l'algoritmo k-means, sia algoritmi di clustering gerarchico, al fine di scoprire connessioni gerarchiche tra i gruppi di dati considerati. La seconda tipologia di algoritmi di ML non supervisionato che si intende applicare è basata sulla scoperta di regole associative dai dati. Le regole associative identificano le correlazioni tra gruppi di coppie <attributo-valore>. Ad esempio, la regola:

$$a1 == v1 \ \& \ a2 == v2 \ \Rightarrow \ a3 == v3$$

indica che quando gli attributi $a1$ e $a2$ hanno rispettivamente valore $v1$ e $v2$, allora si verifica che l'attributo $a3$ ha valore $v3$. Gli algoritmi che scoprono regole associative dai dati (il più noto è Apriori) calcolano anche due valori statistici, detti supporto e confidenza che misurano, rispettivamente, la generalità e l'affidabilità della regola.

3 SINTESI

Il progetto punta alla realizzazione di un sistema di supporto alle decisioni, basato su strumenti di intelligenza artificiale, capace di migliorare il dialogo e accorciare le distanze tra individui e aziende o amministrazioni, nel contesto della tutela del credito e per il sostegno della spesa corrente delle famiglie.

Si vuole in tal modo rispondere a due esigenze, strettamente correlate pur se apparentemente contrastanti: da un lato è necessario ridurre il costo per gli individui (in veste di utenti o di consumatori) di eventuali disavanzi tra entrate correnti e consumi, anche occasionali, dall'altro si deve favorire l'accesso dei cittadini al credito, che deve quindi essere tutelato, per finanziare al tempo stesso il consumo delle famiglie e la produzione nazionale di beni e servizi.

In Italia, l'aumento del costo di beni e servizi e la perdita occupazionale hanno eroso nel corso degli anni le capacità di autofinanziamento delle famiglie, rendendo strutturale il ricorso al prestito e alto il rischio di sovraindebitamento. Nel contesto di questa fragilità, i cittadini risentono in modo particolarmente negativo delle disuguaglianze di posizione spesso generate da condotte asimmetriche da parte di aziende e pubbliche amministrazioni, come dell'eventualità di fatturazioni di congrui elevati pluriennali.

Settori particolarmente interessati da questi fenomeni sono le telecomunicazioni, i servizi idrici integrati, i servizi energetici, il fisco e la pubblica amministrazione e i servizi bancari, tra cui in particolare quelli di credito al consumo.

Parallelamente a quanto sopra, proprio per rispondere all'erosione del potere d'acquisto delle famiglie, le aziende produttrici di beni e servizi hanno stabilito un ricorso divenuto ormai strutturale a sistemi finanziari capaci di erogare crediti al consumo, senza i quali – oggi – si riscontrerebbe una contrazione del mercato tale da rendere insostenibile l'attuale sviluppo economico / occupazionale.

A fronte di queste difficoltà, non è sufficiente rivolgere accuse a un'etica commerciale troppo spregiudicata o a una pubblicità ingannevole o a un atteggiamento superficiale del consumatore. E' necessario invece rimuovere gli ostacoli attualmente presenti al corretto funzionamento del sistema, che possono molto spesso ricondursi a errori di comunicazione nello svolgimento del rapporto creditizio, spesso acuiti dalle disuguaglianze informative proprio del nostro territorio.

Per rispondere a queste esigenze, il progetto punta allo sviluppo di una particolare applicazione di IA, che, partendo dal know how attuale delle aziende di recupero crediti, sviluppi metodi di ottimizzazione dei rapporti creditizi e migliori il dialogo tra le parti.

A tal fine, il piano di lavoro si articola in tre macro-aree, ciascuna comprendente tre obiettivi realizzativi e sviluppata da uno dei partner. La denominazione delle aree evoca il concetto di dialogo, tra cittadini / utenti / consumatori, da un lato, e agenti demandati da aziende e amministrazioni, dall'altro, per il recupero dei crediti. La prima, sviluppata dall'Università di Messina, riguarda il "sistema interlocutore", la seconda, facente capo a Expert-System, il "sistema dialogo" e la terza, di Fire, il "sistema scelte". Si intende infatti, in primo luogo nell'ambito del "sistema interlocutore", definire una classificazione dedicata, applicabile nel dominio del credito e del consumo, ovvero – in particolare – un profilo multidimensionale, capace di rappresentare il cittadino / utente / consumatore in maniera completa e non ambigua per gli scopi del progetto. In secondo luogo, nel "sistema dialogo", si vuole realizzare una rappresentazione della conoscenza dedicata, con un'ontologia linguistica adatta, valida all'interno degli scopi del progetto precedentemente menzionati, ottenibile partendo dall'analisi di dati storici mediante TIC quali: Automatic Speech Recognition (ASR), Text Mining, Natural Language Processing (NLP), categorizzazione automatica, Named Entity Recognition (NER). Infine, con il "sistema scelte" si intende capitalizzare e formalizzare l'esperienza del proponente Fire nella gestione delle pratiche creditizie, allo scopo di perseguire un'ottimizzazione semi-automatica dei percorsi di recupero.

L'output finale del progetto sarà un sistema di supporto alle decisioni, capace di assistere gli interlocutori nello sviluppo delle pratiche di recupero crediti e salvaguardare in tal modo la capacità creditizia dei cittadini e la liquidità delle imprese e delle amministrazioni.

4 FINALITÀ

Lo scenario di riferimento del progetto è quello del rapporto tra economia reale e finanziaria, con particolare riferimento al mercato dei crediti e all’impatto che questo mercato ha sullo sviluppo economico.

A seguito della crisi finanziaria generata dai mutui residenziali statunitensi (2008) e della crisi del debito sovrano dell’Area euro (2011) si è verificata una caduta generalizzata degli investimenti e una contrazione dei consumi. Il credito concesso in Italia alle famiglie ha conseguentemente subito forti contrazioni, giungendo – tra ottobre 2011 e novembre 2013 – al punto di minimo ⁽¹⁾. Dal 2014 l’andamento si è invertito, superando, nel 2017, quota 20 500 €/famiglia ⁽²⁾. Nello stesso anno, gli impieghi bancari per l’acquisto di abitazioni sono aumentati dell’1.9 %, mentre il credito al consumo – concesso da banche e finanziarie – è cresciuto dell’8.3 %. Secondo l’Osservatorio Compass (gruppo Mediobanca), il totale del credito al consumo erogato è salito a 19 miliardi, con l’80% degli acquisti che non sarebbero stati effettuati in assenza di tale credito ⁽³⁾. I dati attualmente pubblicati dalla Banca d’Italia per il 2018 confermano quelli del 2017 ⁽⁴⁾.

La crisi finanziaria e il deterioramento del quadro economico hanno ridotto nel corso degli anni le entrate delle famiglie, per cui il reddito medio equivalente è sceso dal 2006 al 2014 per risalire solo nel 2016, a circa 18 600 €, il 3.5% in più rispetto al 2015 ^(5, 6).

Questo recente andamento favorevole ha fatto crescere il numero delle famiglie che sono riuscite a risparmiare parte del reddito nel corso dell’anno (dal 27 al 33 %), ma, tra quelle appartenenti al 30% più povero, è anche cresciuto il numero di quelle che hanno dovuto ricorrere ai risparmi o all’indebitamento per finanziare la propria spesa. **Il rapporto ISTAT sulla situazione sociale italiana nel 2017 mostra che, sui quasi 26 milioni di famiglie italiane, circa 1/4 è disagiata.**

Occorre notare (v. figura 4.1) che il 70% delle famiglie ha una ricchezza non superiore a 200 000 €; se ci si limita ad analizzare questa popolazione, si vede che essa contiene una percentuale di famiglie indebitate superiore a quella delle fasce più ricche, con percentuali anche superiori al 25%, e valori medi delle passività dello stesso ordine di grandezza di famiglie molto più ricche (dal percentile 30 al percentile 95, ovvero da 0 a 680 000 € di ricchezza netta, il valore medio delle passività è dell’ordine di 50 000 €).

Nel contesto delle crisi finanziarie sopra accennate, il costo di beni e servizi è progressivamente aumentato rispetto ai salari. Nel periodo 2000-2013 l’indice dei prezzi al consumo per le famiglie di operai e impiegati è sceso, mediamente, del 2.5%/anno, stabilizzandosi tra il 2013 e il 2017 con una perdita complessiva del potere di acquisto rispetto al 2000 del 32.5% (elaborazione da dati ISTAT, https://www.istat.it/it/files//2011/06/coefficienti_annuali_1861_2017.pdf).

Se si considera che l’indebitamento calcolato statisticamente della figura 4.1 riguarda l’accensione di mutui per l’acquisto di una abitazione, i prestiti personali, i prestiti contro la cessione dello stipendio e le aperture di credito in conto corrente (in genere forme di credito al consumo), mentre trascura le passività legate a imposte e bollette non pagate, si capisce come la pressione sui crediti delle famiglie meno abbienti sia molto alta e forse prossima a livelli critici.

Nel contesto di questa fragilità, i cittadini risentono in modo particolarmente negativo delle disegualianze di posizione spesso generate da condotte asimmetriche da parte di aziende e pubbliche ammi-

¹Stefania Pozzuoli, “L’evoluzione del credito alle società non finanziarie e alle famiglie: un’analisi empirica per l’Italia” Ministero dell’Economia e delle Finanze Working Papers N°2 - Febbraio 2018

²Dati CGIA 31/12/17, fonte ANSA

³Vittoria Puledda “Credito al consumo, il 2017 è stato un anno boom” La Repubblica, 10/05/2018.

⁴Banca d’Italia “Indagine sul credito bancario nell’area dell’euro. Principali risultati per le banche italiane. 3° trimestre del 2018 e prospettive per il 4° trimestre del 2018”, ottobre 2018.

⁵Nicola Barone “Si ferma la caduta del reddito medio ma aumenta la povertà” Il Sole 24 Ore, 12/03/2018.

⁶Banca d’Italia “Indagine sui bilanci delle famiglie italiane” 12/03/2018.

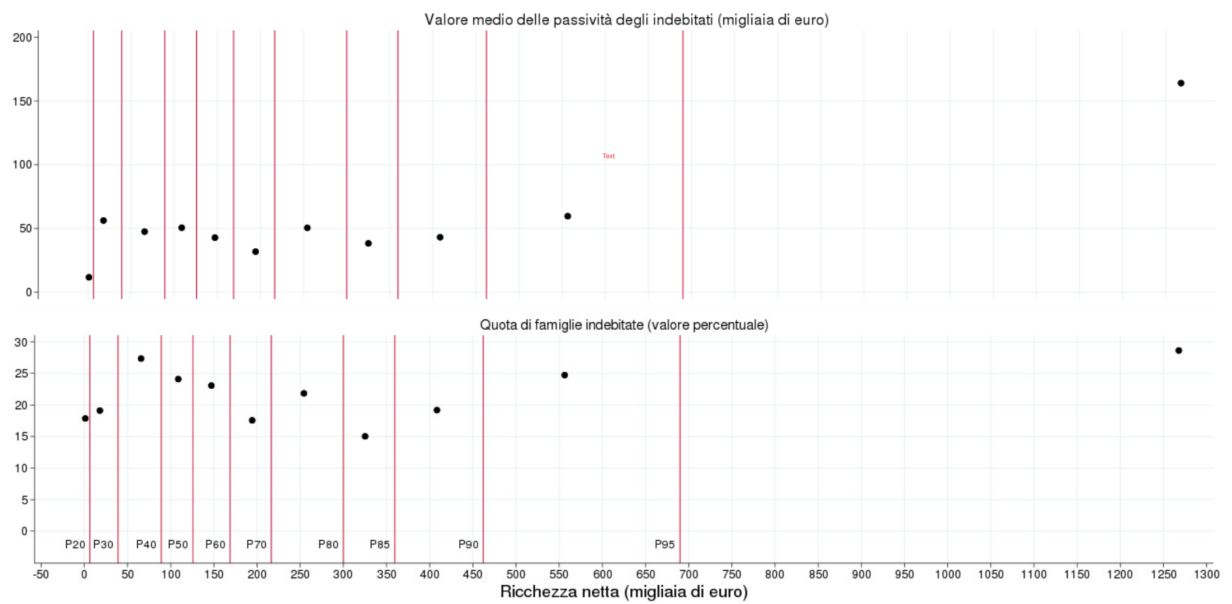


Figura 4.1: Passività finanziarie delle famiglie nel 2016 (migliaia di euro e valori percentuali). “Fonte: Elaborazioni sull’archivio storico dell’Indagine sui bilanci delle famiglie italiane, versione 10.0. Nota: i punti indicano valori medi e percentuali per le famiglie con ricchezza netta appartenenti agli intervalli individuati dalle linee verticali che indicano alcuni selezionati percentili della distribuzione della ricchezza netta (il percentile Px indica il valore della ricchezza netta al di sotto del quale ricade la percentuale x di famiglie che hanno ricchezza non superiore a tale valore; ad esempio, il 70 per cento delle famiglie ha una ricchezza non superiore al percentile P70 netta mentre il restante 30 per cento ha una ricchezza maggiore)” [Banca d’Italia op. cit.].

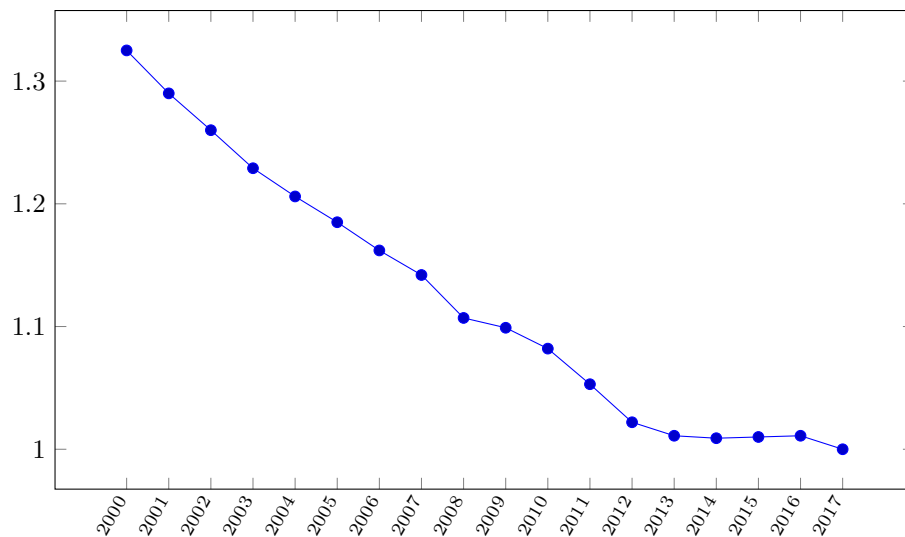


Figura 4.2: Indice nazionale dei prezzi al consumo per le famiglie di operai e impiegati [ISTAT].

nistrazioni, come dell'eventualità di fatturazioni di congruagli elevati pluriennali. Conseguentemente, lo stress finanziario acquisisce una forte componente asociale.

D'altro canto, le famiglie italiane rappresentano circa il 60 % PIL, per cui, nonostante la percentuale di spesa da attribuirsi alle famiglie sottoposte a maggiore pressione creditizia debba essere stimata come una percentuale ridotta di questa (come sopra indicato, si tratta del 70% delle famiglie che possiede il 30% della ricchezza), si può certamente considerare il fenomeno della difficoltà di rimborso dei crediti come una minaccia attuale per l'economia nazionale.

Per ridurre il peso di questa minaccia, non essendo auspicabile né percorribile il ricorso a politiche inflazionistiche o di svalutazione, solo corrette politiche di recupero crediti possono avere un impatto positivo sui livelli occupazionali e preservano il valore reale dei salari ⁽⁷⁾. Dati CNEL UNIREC (Unione nazionale imprese a tutela del credito) dimostrano che nel settore finanziario, che conta poco meno di 640 000 occupati, i posti di lavoro mantenuti grazie al recupero crediti è cresciuto dal 5% del 2010 al 9% del 2012, mentre nei servizi di informazione e comunicazione, su circa 550 000 addetti, questa percentuale è passata dal 19% al 17% ⁽⁸⁾.

Se la tutela del credito assume dunque un ruolo fondamentale per la mitigazione dei rischi correlati alla fragilità attuale di parte della popolazione italiana, occorre anche implementare strategie che rendano sostenibile a lungo termine il meccanismo del credito al consumo come si è delineato negli ultimi anni in altri Paesi.

Una progressiva, generale, deregolamentazione ha reso il credito al consumo un business molto attraente per diverse istituzioni finanziarie operanti nel mondo occidentale. Tanto che ad esempio negli Stati Uniti i media non hanno esitato a definire l'attuale generazione di giovani come "generazione di debitori". In diversi Paesi OCSE, ai giovani è possibile accedere al prestito a consumo e a prestiti ponte in attesa che arrivi il prossimo stipendio non appena raggiungono i 18 anni d'età. Questi prestiti tipicamente hanno tassi relativamente alti e devono essere ripagati in un arco di tempo breve, soprattutto se confrontati con mutui oppure prestiti per studenti. Prestiti ponte sono tipicamente accessibili anche da coloro che non hanno un impiego stabile o asset per garantire tali debiti e/o ottenerli a tassi più bassi.

Il mercato dei prestiti al consumo ha raggiunto dimensioni ragguardevoli. Ad esempio, nel Regno Unito, i dati dell'Office of Fair Trading riportano una crescita da 900 milioni di sterline del 2008 a 2 miliardi del 2012, per passare agli oltre 3 miliardi attuali. Ci sono oltre 100 compagnie nel Regno Unito che offrono tali prodotti per una occupazione di oltre 6000 unità. In generale, tali prestiti possono essere visti in maniera positiva in quanto consentono agli individui in situazione di temporanea difficoltà la possibilità di coprire costi ed onorare i propri impegni. Inoltre, visto che la quota di popolazione che maggiormente utilizza tali strumenti appartiene alla fascia giovane di età compresa tra 26 e 45 anni (53,4%), teoricamente si potrebbe pensare che la popolazione in fascia di età produttiva non abbia problemi a ripagare tali debiti.

Tuttavia, queste nuove opportunità di ottenere finanziamenti in maniera veloce sono state considerate come una delle cause maggiori di stress finanziario per le giovani generazioni, e hanno spesso determinato default e procedure esecutive per la soddisfazione del credito. **Default e insolvenza stanno diventando fenomeni comuni, soprattutto tra le giovanissime generazioni in età tra 18 e 25 anni che sono ancora nella fase di apprendimento di concetti quali management e indipendenza finanziaria.**

Soprattutto all'indomani della crisi mondiale del 2008, i cui effetti macroeconomici non sono ancora stati del tutto assorbiti, il tasso di disoccupazione è cresciuto sensibilmente e i prestiti ponte e al consumo si sono trasformati da strumenti finalizzati ad aiutare gli individui in stato di temporanea difficoltà a strumenti finalizzati a soddisfare esigenze di liquidità giornaliera. Inoltre, la facilità con la quale spesso questi prestiti vengono erogati anche a individui con basse capacità di reddito e precedenti episodi di default, pongono quesiti di irresponsabilità da parte delle istituzioni creditizie che li erogano. Sempre per citare dati relativi al Regno Unito, la relazione del Governo riporta che negli ultimi 5 anni il 52% dei clienti dei prestiti ponte ha avuto problemi di soddisfazione del debito, il 38% dei clienti ha un credit rating molto basso, l'11% era stato oggetto di procedure esecutive. Le statistiche mostrano anche che il 28% dei prestiti erogati nel 2011 erano stati rinnovati almeno una volta ed il 5% per più di quattro volte, incidendo per più del 50% dei ricavi delle agenzie erogatrici del credito. Da un lato, si ribadisce il

⁷Stefano Gatti, Manuela Geranio "Recupero crediti, attività economica ed effetti occupazionali" VI Rapporto annuale UNIREC, Roma 12/05/2016.

⁸Stefano Gatti, Manuela Geranio op. cit.

ruolo positivo della presenza di tali strumenti, nella misura in cui aiutano a risolvere problemi di liquidità temporanei e di breve periodo. Dall'altro, è lecito chiedersi se un accesso più facile al credito non inneschi una spirale di lungo termine che porta a situazioni di stress finanziario.

A livello internazionale, le conseguenze delle situazioni di stress finanziario da parte delle giovani generazioni variano a seconda della legislazione vigente in materia di insolvenza. Ad esempio, nei Paesi del Nord Europa dove la legislazione è piuttosto incline a proteggere il creditore, diversi studi hanno documentato che situazioni di insolvenza si sono accompagnate a problemi psicologici, ansia, depressione, vergogna e tendenze suicide. Negli Stati Uniti, dove la possibilità di ricominciare una vita normale dopo la bancarotta individuale è molto più semplice, questi effetti sono notati in maniera meno diffusa. Inoltre, altri studi hanno documentato che la presenza di strumenti di credito al consumo aiuta le giovani generazioni ad accelerare il passaggio ad uno stato di piena maturità, in quanto aumentano il senso di responsabilità e di auto-stima.

Nel Regno Unito è possibile accedere a dati più granulari grazie alla presenza della Wealth and Assets Survey. Dall'analisi delle risposte alla Survey si evince che la maggioranza (61%) di individui ha acquisito almeno un prodotto di credito al consumo e che un individuo su quattro è in uno stato di indebitamento netto. Ordinando gli individui sulla base del rapporto tra debito e reddito, si evince che il top 10% degli individui concentra nelle proprie mani un terzo del debito totale e ha un livello di debito superiore a 2.5 volte il reddito del nucleo familiare. Utilizzando una soglia ragionevole per la definizione di stress finanziario, si evince che circa 2.2 milioni di individui si trovano in questa situazione, un dato significativo a livello macroeconomico. Confrontati al totale della popolazione, gli individui in stato di stress finanziario sono più giovani, con reddito più basso e alto livello del rapporto tra debito e reddito e maggioranza di prodotti finanziari ad alto tasso. Tale rapporto empiricamente è un predittore molto robusto di una futura situazione di stress finanziario. Tanto che le autorità monetarie hanno incluso tale rapporto tra le misure cui le istituzioni finanziarie devono fare riferimento per valutare la sostenibilità del finanziamento richiesto.

Intravedendo problematiche simili nella futura società italiana e tenendo conto delle attuali, il progetto punta alla creazione di nuovi strumenti informatici a supporto del sistema di tutela del credito, che aiutino a ricomporre condizioni di stress finanziario tutelando quanto più possibile ciascuna delle parti in opposizione.

A tal proposito, il XVII Rapporto PiT servizi di Cittadinanzattiva indica come possibili vie per la mitigazione dei conflitti tra individui e servizi bancari e finanziari:

- “Educazione finanziaria. Attivare definitivamente una strategia istituzionale con l'obiettivo di coordinare e guidare le iniziative idonee per far crescere la conoscenza, la consapevolezza, la competenza, la cultura degli italiani in materia economico finanziaria.”
- “Sovra-indebitamento. Far conoscere l'opportunità del Piano del Consumatore e promuovere e costituire presso gli enti locali, in particolar modo, presso i segretariati sociali, gli Organismi di Composizione delle Crisi (OCC).”
- “Credito al consumo. Promuovere un fondo di solidarietà per i consumatori che non sono in grado di poter pagare le rate di prestiti e finanziamenti, così come è stato fatto per i mutui, chiedendo ad istituzioni ed imprese senso di responsabilità.”
- “Digitalizzazione banche. Attraverso la “costruzione” di banca del futuro, rivedere completamente la relazione con il consumatore, a partire dal condividere con lo stesso sia la riduzione dei costi che porterà questo cambiamento che la tempestività nella risoluzione delle problematiche.”

Per rispondere a simili aspettative, all'interno del contesto di riferimento, si ritiene particolarmente utile il ricorso a sistemi di intelligenza artificiale per sviluppare metodi di ottimizzazione dei processi relazionali e di gestione dei rapporti tra le parti.

Il recupero crediti è infatti attualmente un'attività ad alta intensità di risorse umane, nella quale non è facile correlare situazioni ricorrenti, se non attraverso l'esperienza acquisita dagli operatori, né è facile individuare percorsi ottimali entro i quali eventualmente inscrivere rapporti simili. Se si aggiunge alla complessità tecnica delle diverse pratiche di credito quella umana legata ai rapporti con un interlocutore

spesso in gravi condizioni di stress o di rifiuto di un rapporto che considera asimmetrico, si capisce quanto possa essere vantaggioso il supporto di un sistema di supporto alle decisioni.

Il contesto tecnologico di riferimento è allora quello delle interfacce conversazionali, agenti software in grado di interagire in linguaggio naturale (scritto o parlato) con un interlocutore umano, frontiera attuale nel mondo dei customer service. Si intende tuttavia andare oltre il livello delle attuali chatbot.

Gli studi eseguiti in questo contesto mostrano che i proprietari di prodotti tecnologici sono sempre meno soddisfatti dai customer service dedicati in quanto questi stentano a integrare nei loro processi la capacità di comprendere il reale problema del cliente e il modo in cui egli vuole risolverlo. Nonostante questo, il 63% degli interlocutori preferiva, già nel 2015, un contatto digitale, probabilmente in quanto più rapido ⁽⁹⁾.

Secondo uno studio dell'Osservatorio Artificial Intelligence della School of Management del Politecnico di Milano, il 56% di 721 imprese analizzate ha avviato progetti di IA, di cui il 25% su virtual assistant / chatbot ⁽¹⁰⁾. Il fatto che questi strumenti siano prevalentemente impiegati come primo livello di contatto con il cliente che si rivolge al customer service aziendale conferma il ruolo di frontiera rivestito da impieghi più avanzati, attualmente riservati a operatori umani.

Traslando il problema dal contesto dei prodotti hi-tech a quello in esame, si intuisce come il bisogno di comprensione del punto di vista dell'interlocutore sia ancora più importante, tanto più dato che l'obiettivo del rapporto è quello di ricomporre una condizione di stress che può avere gravi conseguenze sociali. Lo sviluppo di applicazioni di IA capaci di interagire in un simile contesto comporta quindi necessariamente importanti azioni di ricerca.

Riassumendo:

- La tutela del credito è una funzione economica fondamentale sia nel contesto attuale italiano che in quello prossimo futuro; il suo peso è destinato a crescere e ad assumere valore sociale.
- Nello svolgimento di azioni di recupero è necessario tenere conto del contesto, delle esigenze e delle aspettative degli individui.
- Nel rapporto con gli interlocutori occorre capacità di integrazione di dati eterogenei, comprensione dei problemi, memoria del pregresso, rapidità di analisi e intraprendenza per la proposta di soluzioni.
- Le questioni precedenti possono essere risolte attraverso un sistema avanzato, non ancora esistente, ma attualmente realizzabile, di IA.

Finalità del progetto è dunque lo sviluppo di una particolare applicazione di IA, che, partendo dal know how attuale delle aziende di recupero crediti, sviluppi metodi di ottimizzazione dei rapporti creditizi e migliori il dialogo tra le parti.

Nel logo del progetto, si è voluto anche evidenziare come in MIRC.0 il dialogo tra le due parti, interlocutore e azienda di recupero del credito, possa comporsi (i due triangoli in figura) e concorrere a una soluzione soddisfacente per entrambe all'interno di un'unica forma (il poligono).

⁹Blake Morgan "The State Of Customer Experience Study 2017" Forbes 31/10/17.

¹⁰Maria Teresa Della Mura "Migliorare il customer service con l'intelligenza artificiale" AI4business 07/11/18.

5 OBIETTIVO FINALE DEL PROGETTO

5.1. Caratteristiche e prestazioni

Caratteristiche Per raggiungere la finalità sopra indicata, l'obiettivo finale del progetto è la realizzazione di un sistema di IA capace di:

1. Acquisire esperienza dai dati storici sulle pratiche di recupero crediti;
2. Acquisire esperienza durante lo svolgimento dei propri compiti sulle pratiche attuali;
3. Identificare l'interlocutore secondo un'apposita classificazione, che tenga conto – tra l'altro – delle condizioni attuali e future di stress finanziario che possono averlo colpito;
4. Comprendere il problema specifico posto all'interlocutore dalla pratica in corso;
5. Individuare il migliore percorso, integrando informazioni provenienti da tutte le fonti necessarie allo scopo, per ottenere la soluzione della controversia, secondo il profilo dell'interlocutore e a partire da una qualunque condizione a priori – dove per “migliore” si intende un percorso che tenga conto in modo responsabile del peso sociale dell'operazione;
6. Scambiare informazioni con l'interlocutore in modo da procedere secondo percorsi ottimali, via via aggiornati sulla base di queste informazioni, fino a risolvere la pratica.

Il sistema dovrà essere inserito all'interno degli attuali processi produttivi del proponente Fire, interagendo con gli operatori (dai phone collectors agli analisti, fino al top management) e con gli interlocutori (debitori), realizzando un'innovazione di processo (il sistema migliora i processi produttivi attuali) attraverso una fondamentale innovazione di prodotto (non esistono in commercio sistemi del genere).

Prestazione 1: gestione del rischio Attualmente, presso Fire come in altre imprese del settore, i modelli di ottimizzazione della gestione delle pratiche di recupero crediti e di definizione del profilo di rischio del debitore utilizzano metodi classici, basati su strumenti di regressione lineare o di associazione statistica. L'idea di questo progetto è di sfruttare le recenti tecniche di analisi e di classificazione utilizzando strumenti di machine learning (ML). ML è particolarmente adatto ad analizzare database anche non strutturati, multidimensionali e molto grandi (come quelli di Fire) per cui l'aspettativa è di migliorare notevolmente la capacità predittiva del rischio e di rendimento atteso da ogni pratica e, al contempo, migliorare l'identificazione della modalità di recupero ottimale anche per l'interlocutore. Questo miglioramento può consentire di individuare le pratiche a più alta probabilità di soddisfazione del credito e concentrare l'utilizzo di risorse umane su queste ultime, permettendo all'impresa di fare leva sulla esistente forza lavoro per incrementare l'attività di gestione. Una migliore capacità predittiva del rischio può anche consentire di spingere al di là dei livelli oggi accettabili la trattativa con interlocutori ritenuti affidabili (per esempio, individui sottoposti a momentanea pressione economica comprovata dall'incrocio di dati).

Prestazione 2: sistema di supporto alle decisioni All'interno dell'iter di gestione di ogni pratica di credito, le fasi iniziali di phone servicing sono quelle a maggior intensità di forza lavoro; esse pertanto incidono maggiormente sui margini operativi e profitti dell'impresa e sui rapporti con gli interlocutori. L'operatore telefonico deve spesso interagire con il livello gerarchico superiore e questo comporta tempi che potrebbero essere ridotti grazie all'interazione con un sistema esperto. Inoltre questa interazione tra

operatori umani genera, nel contesto attuale, rischi di inefficienza, perdita di informazione, riduzione delle opzioni, che possono essere evitati col ricorso a strumenti informatici. Una seconda idea del progetto è dunque quella di realizzare un **sistema esperto interattivo**. Si ritiene che in un primo tempo il sistema possa limitarsi ad ascoltare le conversazioni e agire solo attraverso gli operatori umani, mediante un'interfaccia appositamente progettata, lasciando agli stessi la possibilità di decidere sul proseguimento del rapporto con l'interlocutore, per passare in un secondo tempo – gradualmente e per i passaggi più semplici – a un'interazione diretta con gli interlocutori mediante un'interfaccia robotica.

Prestazione 3: interfaccia conversazionale La terza idea del progetto propone di utilizzare le recenti trasformazioni tecnologiche nel settore della robotica della comunicazione per cercare di automatizzare e robotizzare quanto più possibile la fase di phone servicing.

Innovazione di prodotto Il progetto punta alla creazione di un nuovo sistema di supporto alle decisioni (che opera in modo totalmente diverso da quanto normalmente fatto dai software di credit scoring), avente tre funzioni principali: a) tenere conto di un profilo complesso dell'interlocutore e della pratica che lo riguarda allo scopo di venire incontro alle sue reali esigenze; b) suggerire percorsi ottimali, per ciascuna pratica di credito, e valutare la probabilità di soddisfazione del credito stesso, corredando questa informazione con la durata prevista dell'intera gestione della pratica in funzione dei percorsi alternativi; c) **pervenire alla determinazione di un costo di ciascuna pratica** (o di un lotto di pratiche, tenendo presente le correlazioni che potenzialmente ne riducono il rischio). **L'anticipazione del costo consente sia di rivelare all'impresa che tipo di margini applicare per gestire la singola oppure il lotto di pratiche, sia di rivelare un fair value, utile nel caso la Fire decidesse di assumere in proprio il rischio di non evasione della stessa.**

Innovazione di processo derivata dal nuovo prodotto Dal punto di vista dell'innovazione di processo, il progetto giunge a definire un nuovo modello di business per la gestione del credito, che sposta il focus da un approccio human driven a uno sempre più data driven. Per l'interlocutore questo spostamento può condurre all'ampliamento delle garanzie che alcune reali condizioni di sofferenza possano essere recepite, visto che il sistema può avere accesso a una mole molto più importante di informazioni dell'operatore umano e l'interlocutore e la pratica che lo riguarda possono essere profilati in modo molto più completo e complesso. Per l'impresa, i cui profitti si definiscono in termini di margini sul ricavo, un focus data driven permette di spalmare la forza lavoro esistente su un numero maggiore di pratiche e di focalizzarla su quelle a più alta probabilità di successo, garantendo un impatto positivo sui profitti.

Il nome del sistema che si vuole produrre è MIRC.0, per indicare la versione 0 di una Macchina Intelligente Recupero Crediti. L'etimologia del nome Mirco è appropriata in quanto evoca qualcuno che porta la pace.

5.2. Principali problematiche tecnico-scientifiche e tecnologiche

Per descrivere le principali problematiche tecnico-scientifiche e tecnologiche da affrontare si può fare riferimento alla lista delle prestazioni che si intende realizzare. Relativamente ai punti 1 e 2 del precedente elenco (§5.1), le questioni sono:

1. **Comprensione e trascrizione automatica di file audio con diversi parlatori, possibile utilizzo di dialetti e qualità del segnale ridotta;**
2. **Comprensione ed estrazione automatica di informazioni da file di testo generati dal riconoscimento di file audio (ASR)**
3. **Comprensione ed estrazione automatica da diverse tipologie di documenti (es. pdf, tabelle, xls,...) di informazioni espresse nei diversi modi possibili, tipici del linguaggio di tutti i giorni, che meglio rappresentino sinteticamente la conoscenza del dominio.**
4. **Identificazione ed estrazione delle categorie ed entità utili per una precisa descrizione del dominio di riferimento (metadati semantici)**

5. Identificazione delle più promettenti combinazioni di metadati (pattern) di tipo semantico e quantitativo per l'identificazione di precise tipologie di utenti (clustering, classificazione automatica).

Relativamente al punto 3 della lista del §5.1, si tratta di:

6. Valutare l'impiego di un wordnet generico, come ItalWordNet, e realizzare – se necessario – un wordnet terminologico relativo al dominio in esame, collegato al precedente (come fatto, per esempio, dall'Istituto di Linguistica Computazionale del CNR nel caso della navigazione e del trasporto marittimo¹);
7. Sviluppare un modello concettuale dell'interlocutore, che ne definisca un profilo multidimensionale, contenente informazioni specifiche a priori (quali: condizioni economiche, sociali, psico-caratteriali, storia creditizia, eventi recenti familiari e lavorativi) e contingenti (motivazioni del debito, tipologia dello stesso, ...);
8. Identificare e implementare metodi di acquisizione dei dati per alimentare il suddetto profilo, secondo i limiti previsti dalla Legge, sia sulla base dei dialoghi diretti o registrati con l'interlocutore (v. punto 5 della lista precedente) che attraverso altri data repository.

Per quanto riguarda il punto 4 e 6 §5.1, si tratta di sviluppare un sistema intelligente capace di comunicare in modo produttivo con esseri umani sfruttando le capacità di comprensione del linguaggio le cui problematiche sono state precedentemente enumerate (prima lista di questa sezione). L'argomento è di grande attualità, con nuovi risultati continuamente emergenti, per cui si intende:

9. Identificare le migliori tecniche di apprendimento automatico, sulla base di una estensiva sperimentazione e test;
 10. Valutare e garantire la scalabilità delle tecnologie identificate, al crescere della mole di dati.
- Infine, il punto 5 §5.1 si inserisce all'interno delle seguenti problematiche:
11. Informatizzazione delle procedure, comprendente l'interfaccia con basi di dati sulla regolamentazione in vigore nei diversi contesti di interazione;
 12. Definizione di protocolli e sistemi per la qualità e la sicurezza dei dati (data quality e data security);
 13. Ingegnerizzazione delle conoscenze del personale Fire;
 14. Definizione della metodo di misura del raggiungimento degli obiettivi.

I punti della lista precedente sono ordinati in modo crescente secondo il peso che in essi giocano scelte di natura socio-politica più che scientifico-tecnologica. In particolare, la definizione dell'obiettivo da raggiungere, dovendo comporre questioni strettamente creditizie con altre di natura psicologica e sociale, richiederà analisi a posteriori sul tipo di risultato raggiungibile, che dovranno essere organizzate attraverso gruppi di controllo.

5.3. Soluzioni tecnologiche previste

Le soluzioni tecnologiche previste afferiscono prevalentemente al mondo dell'IA, ma altre tecnologie ICT sono convocate nel progetto. Possiamo fare riferimento alle problematiche precedentemente enumerate per indicare le soluzioni previste.

Relativamente ai punti da 1 a 5, si prevede di applicare tecnologie di Natural Language Processing (NLP), più in dettaglio:

- Entity Extraction (riconoscimento di entità presenti nel testo);
- Categorizzazione automatica (classificazione secondo specifiche tassonomie di dominio);

¹Rita Marinelli, Adriana Roventini, Alessandro Enea "Building a Maritime Domain Lexicon: a Few Considerations on the Database Structure and the Semantic Coding" ILC CNR 2004.

MIRC.0

- Writeprint (identificazione di caratteristiche esclusive del testo);
- Mood (calcolo del tono della conversazione).

Relativamente ai punti da 9 e 10, si ricorrerà ad algoritmi di Machine Learning (ML) supervisionato e non supervisionato. Tra le tecniche di ML non supervisionato, si prevede in particolare di esplorare soluzioni basate su:

- Clustering, in particolare tecniche quali k-means e SVM per il clustering gerarchico;
- Data mining, in particolare algoritmi come Apriori per la scoperta di regole associative dai dati.

Relativamente al ML supervisionato, si considereranno i principali algoritmi di classificazione automatica, basati su alberi decisionali e regole classificatorie. Si prevede di valutare l'adozione di un sistema open-source (Weka) o di sistemi di apprendimento automatico probabilistico sviluppati da UNIFE.

Come sopra accennato, altre tecnologie ICT saranno convocate durante il progetto, per la gestione dei protocolli (pulizia, sicurezza o oscuramento dei dati), per la creazione dell'interfaccia utente e per delle interfacce verso le basi esterne di dati.

6 RESPONSABILE DEL PROGETTO

Maurizio Brocato, Ing., PhD, HDR, Professore della “École nationale supérieure d’architecture Paris-Malaquais” (Parigi, Francia) e della “École nationale des ponts et chaussées” (Parigi, Francia), direttore del laboratorio di ricerca “GSA – Géométrie Structure Architecture”.

Nato a Catania il 28/05/1962

Telefono: 346 841 86 67

Mail: maurizio.brocato@me.com

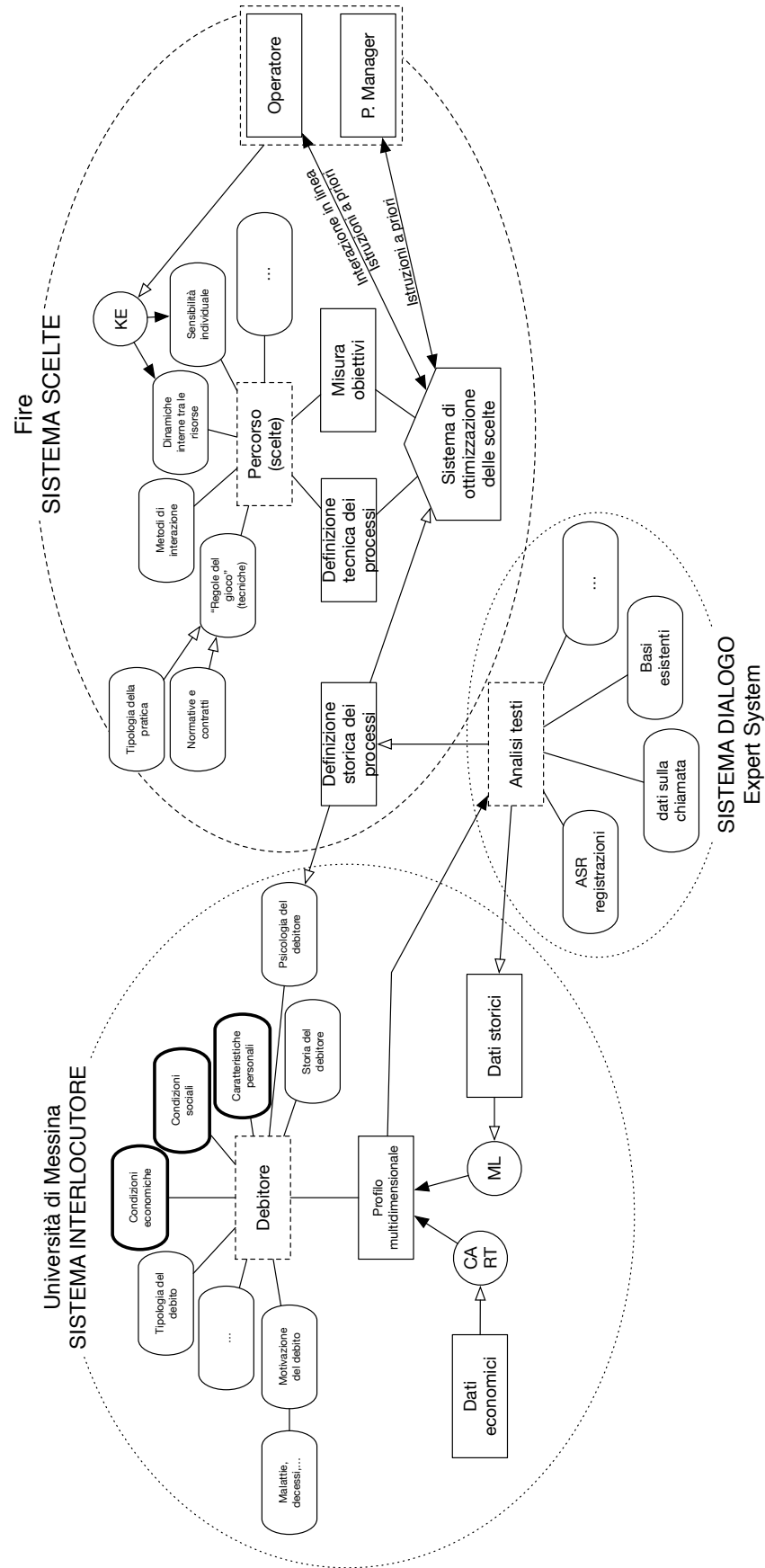
Maurizio Brocato, Ing., PhD, HDR, ha una consolidata esperienza nella gestione di progetti di ricerca e sviluppo, universitari e industriali. Nel corso della carriera, M. Brocato ha partecipato a 22 progetti industriali di R&S, prevalentemente entro i Programmi Quadro Europei, scritto o diretto oltre 50 rapporti di R&S.

Nella sezione 2.1 della III^a Parte di questo documento, insieme a altre informazioni sul profilo del responsabile scientifico, si riporta la lista dei progetti di ricerca finanziata che egli ha diretto negli ultimi dieci anni.

7 OBIETTIVI REALIZZATIVI DEL PROGETTO

OR	Soggetto proponente	Tipologia obiettivo (SS/RI)	Titolo OR
Sistema interlocutore			
OR1	UNIME	RI	Machine learning
OR2	UNIME	SS	Profilo multidimensionale dell'interlocutore
OR3	UNIME	SS	Applicazione, testing e fine tuning degli algoritmi
Sistema dialogo			
OR4	Expert System	RI	Creazione base dati con informazioni automaticamente estratte da contenuti non strutturati
OR5	Expert System	SS	Estrazione dal testo di informazioni per la profilazione puntuale del soggetto analizzato
OR6	Expert System	SS	Pilota del sistema NLP
Sistema scelte			
OR7	Fire	RI	Data base
OR8	Fire	SS	Sistema utente
OR9	Fire	SS	Valutazione del sistema MIRC.0

Il diagramma riportato alla pagina seguente permette un riferimento sintetico delle diverse problematiche trattate all'interno dello schema progettuale.



7.1. OR1 – Machine learning per estrazione di conoscenza dai dati strutturati**DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO****Titolo dell'OR:** Machine learning per estrazione di conoscenza dai dati strutturati**Soggetto proponente:** UNIME**Tipologia (RI/SS):** RI

Tempistica dell'OR:	mese di inizio	1
	mese di fine	34
	<hr/> durata (in mesi) dell'OR	<hr/> 34

Luoghi di svolgimento: Sede UNIFE

Obiettivo: Sviluppare e applicare soluzioni di Machine Learning (ML, per brevità, o apprendimento automatico) per estrarre conoscenza dalla mole di dati raccolti riguardanti sia le pratiche sia gli interlocutori. Questi dati sono in buona parte già strutturati, ma risultano sparsi (con mancanza di valore per alcuni di attributi). Per le pratiche è anche registrato l'esito, con indicazione del valore recuperato rispetto al valore affidato. Obiettivo è la creazione di un sistema di classificazione automatica (prodotto attraverso tecniche di machine learning, partendo dai dati etichettati) che al presentarsi di una nuova pratica (o di un nuovo debitore), sia in grado di restituire una misura prevista del grado di successo nell'affrontare tale pratica.

Parametri che permettono di misurare il raggiungimento dell'obiettivo:

1. Analisi di varie tipologie di dati (es. diverse tipologie di debito, diverse tipologie di pratiche...), con analisi quantitativa mirata a identificare le variabili che si vogliono modellare;
2. Estrazione automatica dai dati (pratiche e/o debitori), con tecniche di apprendimento automatico, di un classificatore in grado di discriminare tra pratiche in base alla percentuale di successo come rapporto tra recuperato su affidato;
3. Identificazione automatica di gruppi omogenei (dai dati relativi alle pratiche e/o agli interlocutori), con tecniche di apprendimento automatico;

Tipo di intervento:

1. Identificazione delle variabili (attributi) rilevanti per le varie tipologie di debitore, di pratiche e tipologia di debito, avvalendosi anche dell'esito dell'analisi quantitativa svolta in OR2 mirata a identificare le variabili che si vogliono modellare;
2. Identificazione dei sistemi di apprendimento automatico (ML) più idonei per identificare similarità (ad es., clustering) e/o per produrre un classificatore automatico da applicare su nuovi casi;

Principali problematiche da affrontare:

1. I dati disponibili risultano sparsi, occorre quindi identificare le variabili (attributi) più rilevanti per le varie tipologie di debitore, di pratiche e tipologia di debito;
2. Identificazione delle migliori tecniche di apprendimento automatico, sulla base di una estensiva sperimentazione e test;

Tempi e modalità di realizzazione dell'intervento: Secondo la lista delle attività riportata

Test da effettuare per misurare i parametri e tempi di realizzazione: In ambito del ML supervisionato, la performance del classificatore automatico appreso si misura principalmente con i valori di accuratezza, precision e recall. Si riportano i concetti utili a tale misurazione nel seguito).

L'accuratezza dà contezza del numero di predizioni corrette rapportato al numero di predizioni totali. La precision indica l'eventuale presenza di falsi positivi. La recall indica l'eventuale presenza di falsi negativi. Nel corso del progetto si procederà con:

1. misurazione della precisione della classificazione automatica delle pratiche, considerando dapprima due sole classi (recuperato nullo, o recuperato non nullo), con obiettivo di precision $> 70\%$ e recall $> 65\%$;
2. successivamente, nel caso di recuperato non nullo, misurazione della precisione nel caso di più di due classi, ove ciascuna classe sia significativa di un intervallo percentuale di recuperato su affidato (ad esempio, entro il 10% di recuperato su affidato, tra il 10 e il 20%, tra il 20 e il 30%, etc...), con obiettivo di precision $> 70\%$ e recall $> 65\%$ per ciascuna classe.

Note descrittive generali: Le attività dell'OR si collocano nel cosiddetto Machine Learning (ML), applicando in particolare tecniche non supervisionate per il Clustering dei dati (di pratiche e/o debitori) e tecniche supervisionate di Classificazione automatica dei dati.

Si applicheranno tecniche di ML supervisionato e non.

Nel ML supervisionato, i dati sono etichettati con valori per una o più classi. Nel caso in esame, ad esempio, ogni pratica registrata può essere etichettata con un valore, discretizzabile come valore percentuale tra 0% e 100%, che indica il recuperato su affidato. E' comunque fondamentale identificare una o più etichette di interesse ai fini della classificazione, e la disponibilità di una base dati in cui ai dati siano associate tali etichette. I dati etichettati sono quindi la base per identificare training set e test set con cui addestrare e misurare le performance del classificatore appreso. Tra le tecniche di ML supervisionato, si applicheranno quelle che apprendono alberi di decisione, o le regole di produzione (che diventano di fatto regole classificatorie).

Gli alberi di decisione sono strutture in cui ogni nodo interno è associato a un test su un attributo e ciascun nodo figlio è associato a un risultato del test. I nodi foglia, invece, sono associati ad una classe o a una distribuzione di probabilità sulle classi. In pratica ogni nodo (radice o intermedio) rappresenta un test su una variabile (ad esempio, attributo $a1==?$) e gli archi uscenti portano a tanti nodi figlio quanti sono i possibili valori della variabile. La costruzione dell'albero avviene in modo automatico dai dati, con algoritmi che individuano la variabile (attributo) più discriminante, e pongono il test su questa come nodo iniziale, iterando poi tale procedimento sulle variabili rimanenti, sino ad arrivare a nodi foglia. Dato un albero decisionale, è possibile classificare un nuovo caso partendo dalla radice, eseguendo il test relativo e scegliendo il figlio associato al risultato del test. Ripetendo questa procedura si arriva ad una foglia: la classe associata alla foglia, o la distribuzione di probabilità sulle classi, rappresenta la classificazione cercata per il nuovo caso.

Gli algoritmi ML costruiscono alberi a partire dal training set in maniera ricorsiva, dividendolo successivamente in insiemi sempre più piccoli scegliendo un test e fermandosi quando l'insieme è omogeneo (ovvero contiene esempi di una sola classe). Il test da scegliere ad ogni passo viene selezionato usando euristiche come il guadagno di entropia o il Gini index. Le regole di produzione invece sono delle implicazioni del tipo *if ... then ...* in cui nella parte *if ...* (antecedente) ci sono test su attributi, e nella parte *then* (conseguente) ci sono classi. Data una lista di queste regole, si può classificare un nuovo caso cercando la prima regola il cui antecedente è verificato per il caso: la classe nel conseguente viene restituita come classe del nuovo caso. Gli algoritmi per apprendere regole di produzione costruiscono una regola alla volta compiendo una ricerca nello spazio delle regole, ordinato secondo la relazione di generalità. Il gruppo di UNIFE ha sviluppato algoritmi di apprendimento di regole di produzione probabilistiche che saranno applicate ai dati disponibili (gli algoritmi sono raccolti al sito ml.uniufe.it). Nel caso di regole probabilistiche, l'algoritmo di ML calcola anche – sempre in base ai dati etichettati – un valore di probabilità che rappresenta il grado di certezza della regola.

Tra le tecniche di ML non supervisionato, il clustering ha l'obiettivo di selezionare e raggruppare elementi omogenei, in base a una misura di distanza o similarità. Questa similarità (e il suo opposto) è

concepita in termini di distanza in uno spazio multidimensionale, ove ogni attributo considerato contribuisce per una dimensione. Diventa quindi fondamentale individuare le variabili (attributi) maggiormente significative per la mole dati a disposizione (di questa attività di occupa UNIME), sia definire la misura della distanza, da decidere insieme agli esperti di Fire, in modo da tenere conto delle caratteristiche dei dati. Si considereranno sia algoritmi quali k-means clustering (un algoritmo iterato che assicura convergenza) dove l'appartenenza a un cluster è determinata sulla base della distanza da un centroide, sia algoritmi di clustering gerarchico, al fine di scoprire connessioni gerarchiche tra i gruppi (di pratiche).

La seconda tipologia di algoritmi di ML non supervisionato che si intende applicare è basata su regole associative. Le regole associative, introdotte in [Agraval,1993] rappresentano correlazioni tra gruppi di coppie <attributo-valore>. Ad esempio, la regola:

$$a_1 == v_1 \ \& \ a_2 == v_2 \ \Rightarrow \ a_3 = v_3$$

indica che quando gli attributi a_1 e a_2 hanno rispettivamente valore v_1 e v_2 , allora si verifica che l'attributo a_3 ha valore v_3 . Gli algoritmi che scoprono regole associative dai dati (il più noto è Apriori) calcolano anche due valori statistici da usarsi come metrica, detti supporto e confidenza. La prima metrica (supporto) misura la generalità della regola, ovvero la numerosità dei casi a cui si applica (la numerosità dei casi in cui sono veri sia l'antecedente sia il conseguente). La seconda (confidenza) misura l'affidabilità della regola, ovvero quanto è probabile il conseguente se è vero l'antecedente (si misura facendo il rapporto tra il numero dei casi in cui sono veri l'antecedente e il conseguente e il numero dei casi in cui è vero solo l'antecedente). In pratica, il supporto indica la percentuale di record/transazioni nella base dati esaminata per cui la regola scoperta è verificata. La confidenza indica la percentuale di record/transazioni per cui quando è vera la premessa della regola è vera anche la conseguenza. Ad esempio, un supporto pari a 0,02 indica che la regola è vera per solo il 2% dei record/transazioni; una confidenza pari a 0,8 indica che nell'80% dei record/transazioni in cui è vero l'antecedente della regola è vero anche il conseguente (il che significa che nel 20% dei casi la regola è smentita). In un'applicazione reale, quale quella in esame, si estrarranno le regole più significative, ovvero quelle con supporto elevato (almeno 50%) e confidenza almeno del 70%, affinché l'analisi produca conoscenza statisticamente significativa. Nel corso del primo anno si applicheranno tecniche di ML non supervisionato (clustering in particolare), al fine di identificare gruppi omogenei di pratiche e/o debitori, e analizzarne le caratteristiche salienti.

Si applicheranno inoltre tecniche di ML supervisionato, per classificare le pratiche e/o i debitori, sulla base di etichettature già disponibili (ad esempio, percentuale di recuperato su affidato). Si considereranno, in particolare, tecniche supervisionate basate su alberi decisionali, e regole di produzione anche probabilistiche, utilizzando software open-source o sviluppati ad hoc dal partner responsabile (UNIME) e/o dal sub-contraente (UNIFE). Se ne valuteranno le performance secondo le metriche individuate.

L'esito di questa prima fase di attività sarà fornito al "sistema interlocutore", in capo ad Expert Systems, e al "sistema scelte" in capo a Fire.

Analogamente, l'esito del primo e secondo anno di attività di OR1 sarà utile per la realizzazione di un classificatore più fine, che tenga conto dell'esito dell'analisi testuale delle conversazioni registrate.

Nel corso del secondo anno si applicheranno anche tecniche non supervisionate basate sulla scoperta di associazioni (regole associative) dai dati registrati o estratti dalle pratiche (e dalle conversazioni loro pertinenti). Inoltre si raffineranno le tecniche di ML supervisionato, per classificare le pratiche e/o i debitori, sia sulla base di etichettature già disponibili (ad es., percentuale di recuperato su affidato), sia sulla base dell'analisi semantica ottenuta dal "sistema interlocutore". Se ne valuteranno le performance secondo le metriche individuate.

L'esito di questa seconda fase di attività sarà utile al "sistema a regole" in capo a Fire.

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE

OR1.1 Identificazione dei dati, etichettatura dei dati e tipologie di analisi

Obiettivo: Identificare i dati disponibili, le loro etichettature ove disponibili, e le tipologie di analisi automatica da svolgere

Organizzazione: Tale attività sarà fondamentale, per individuare da subito l'insieme di dati (strutturati e non) a disposizione, da utilizzare per l'analisi automatica con tecniche di machine learning.

In tale attività si individueranno inoltre le tipologie di analisi da svolgere sui dati (clustering, classificazione) più utili agli scopi del progetto, per estrarre conoscenza ulteriore dai dati strutturati.

Fire tratta circa un milione di pratiche l'anno. Da questo patrimonio dati saranno creati dataset per lo sviluppo e test delle performance degli algoritmi di machine learning (supervisionato e non).

Parte dei dati sarà usato come training set, parte come test set, e per le fasi di benchmarking e fine tuning delle versioni evolutive degli algoritmi di machine learning supervisionato.

Per la classificazione, saranno identificate più classi (etichettature), individuando dapprima due sole classi, e passando poi a più classi (ad esempio, percentuale di recuperato su affidato, o altre modalità di etichettatura suggerite dagli esperti di Fire).

Per tale attività si prevede l'interazione con il personale di Fire, per identificare le caratteristiche dei dati e le tipologie di analisi più utili.

Tempistica:

mese di inizio	1
mese di fine	24
durata (in mesi) dell'attività	24

Consegne:

1. Prima identificazione dei dati (strutturati) disponibili e loro etichettatura: $T0 + 6$ mesi
2. Seconda identificazione dei dati (strutturati) disponibili e identificazione di più etichettature utili al ML: $T0 + 12$ mesi
3. Identificazione finale dei dati (strutturati, e dati estratti dalle conversazioni) utili al ML: $T0 + 24$ mesi

OR1.2 Algoritmi di ML non supervisionato

Obiettivo: Identificare i più utili algoritmi di ML non supervisionato da applicare ai dati disponibili

Organizzazione: In tale attività si identificheranno gli algoritmi di apprendimento automatico non supervisionato più utili al progetto, anche sulla base della prima identificazione di dati disponibili, e tipologie analisi automatiche (non supervisionate) più utili.

Si considereranno in questa attività i principali algoritmi di clustering, quali k-means e SVM, e di data mining quali Apriori per la scoperta di regole associative.

Per tale attività si prevede di investigare l'adozione di un sistema open-source (Weka) disponibile o di altri strumenti analoghi, individuando quello più adatto al tipo di analisi automatica qui necessaria.

Tempistica:

mese di inizio	4
mese di fine	30
durata (in mesi) dell'attività	27

Consegne:

1. Prima identificazione degli algoritmi/strumenti più utili: $T0 + 9$ mesi
2. Prima applicazione degli algoritmi/strumenti ai dati: $T0 + 18$ mesi
3. Retroazione e seconda applicazione degli algoritmi/strumenti ai dati: $T0 + 30$ mesi

OR1.3 Algoritmi di ML supervisionato

Obiettivo: Identificare i più utili algoritmi di ML supervisionato da applicare ai dati etichettati disponibili

Organizzazione: In tale attività si identificheranno gli algoritmi di apprendimento automatico supervisionato più utili al progetto, anche sulla base della prima identificazione di dati etichettati disponibili, e tipologie analisi automatiche (supervisionate) più utili.

Si considereranno in questa attività i principali algoritmi di classificazione automatica, basati su alberi decisionali e regole classificatorie (anche probabilistiche).

Per tale attività si prevede di investigare sia l'adozione di un sistema open-source (Weka) disponibile o di altri strumenti analoghi, sia l'adozione di sistemi di apprendimento automatico probabilistico sviluppati da UNIFE, individuando quelli più adatti al tipo di analisi automatica qui necessaria.

Tempistica:

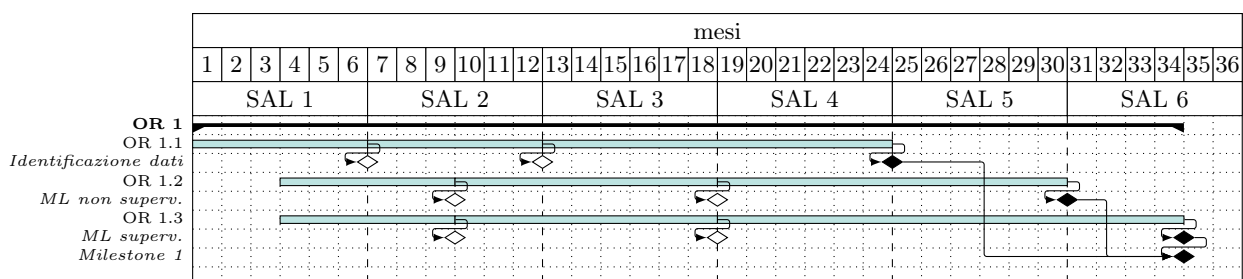
mese di inizio	4
mese di fine	34
durata (in mesi) dell'attività	31

Consegne:

1. Prima identificazione degli algoritmi/strumenti più utili: T0 + 9 mesi
2. Prima applicazione degli algoritmi/strumenti ai dati etichettati: T0 + 18 mesi
3. Retroazione e seconda applicazione degli algoritmi/strumenti ai dati: T0 + 34 mesi

TEMPI DI REALIZZAZIONE

Durata (in mesi): 34

**7.2. OR2 – Profilo multidimensionale dell'interlocutore****DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO**

Titolo dell'OR: Profilo multidimensionale dell'interlocutore

Soggetto proponente: UNIME

Tipologia (RI/SS): RI

Tempistica dell'OR:

mese di inizio	1
mese di fine	36
durata (in mesi) dell'OR	36

Luoghi di svolgimento: Sede UNIME

Obiettivo: Creare un database non strutturato che contenga tutte le informazioni riguardanti il singolo interlocutore. All'interno del totale delle pratiche gestite da FIRE, è frequente il caso che lo stesso interlocutore risulti in diverse singole pratiche affidate. In gergo tecnico l'interlocutore diventa un common factor di queste singole pratiche. Successivamente, sviluppare modelli per testare la capacità predittiva delle varie variabili individuate sulle tipologie di interlocutore.

Parametri che permettono di misurare il raggiungimento dell'obiettivo:

1. Analisi di varie tipologie di fonti (es. report, verbali, database esistenti, pdf, doc, ...);
2. Integrazione delle varie tipologie di fonti;
3. Analisi di casi in cui ci sono dati discordanti sullo stesso tipo di cliente e risoluzione dei conflitti.
4. Analisi della capacità predittiva del modello prescelto.

Tipo di intervento:

1. Integrazione delle variabili (attributi) rilevanti per le varie tipologie di interlocutore;
2. Identificazione di un software che proceda alla integrazione in maniera automatica;
3. Esecuzione di test su campione dei dati esistenti;
4. Progettazione e sviluppo di un dimostratore del sistema di creazione del database.

Principali problematiche da affrontare:

1. Comprensione ed estrazione automatica da diverse tipologie di documenti (es. pdf, tabelle, xls, ...).

Tempi e modalità di realizzazione dell'intervento: Secondo la lista delle attività di seguito riportata.

Test da effettuare per misurare i parametri e tempi di realizzazione: L'accuratezza dei risultati si misura principalmente con la duplicazione degli identificativi dei singoli debitori e con la capacità predittiva del modello individuato.

Il primo indica la correttezza dell'algoritmo di imputazione dei dati. Il secondo indica la bontà del modello. Nel corso del progetto si procederà con:

1. Misurazione delle eventuali duplicazioni di file identificativi;
2. Risoluzione di conflitti tra diverse fonti di dati;
3. Verifica della capacità predittiva dei modelli individuati.

Note descrittive generali: Le attività in questo OR sono di duplice natura. Il primo step consiste in un esercizio di tipo prevalentemente informatico, in quanto consiste nella creazione di un database "madre" che integri le diverse informazioni disponibili a proposito dei vari debitori e delle modalità di espletamento delle diverse pratiche. E' un esercizio preliminare necessario, in quanto assembla in modo organico ancorché non strutturato tutte le informazioni disponibili riguardanti un singolo interlocutore, nonché la loro evoluzione storica. Da questo database granulare sarà poi possibile procedere alla creazione di gruppi, ordinati secondo diverse modalità, che andranno a costituire un input prezioso per l'analisi e lo scoring della singola pratica affidata. La creazione del database partirà con l'inizio del progetto e sarà poi integrata dalle informazioni fornite da Expert System attraverso l'analisi testuale delle conversazioni telefoniche.

Lo step successivo consiste nel formulare modelli predittivi di probabilità di default. Il task si adatta ad essere formalizzato come una procedura di supervised learning, una delle tecniche più utilizzate in letteratura. In questo framework, il modello "impara" da una serie di input/output dati storici, dove gli input sono le caratteristiche dell'interlocutore e / o della pratica e l'output è l'esito della pratica stessa. Il supervised learning può formularsi come un problema di regressione se l'output è continuo, oppure come un problema di classificazione se l'output è discreto. Con i nostri dati analizzeremo entrambi i casi.

Una volta ottenuti i dati input / output, il modello deve trovare una funzione che mappa il vettore di input nell'output osservato. Una possibilità è data dal "memorizzare" tutti i precedenti valori di input output e basare la funzione sull'intero campione. Nonostante così la mappatura sia perfetta, il modello avrà probabilmente una bassa capacità predittiva se gli input sono differenti da quelli riscontrati nel campione utilizzato per "allenare" il modello, oppure se i dati output sono "noisy" (com'è il caso dei dati finanziari). Quindi la sfida consiste nel cercare una funzione che generalizzi la sua applicabilità oltre il campione utilizzato per "allenare" il modello.

Uno degli output del nostro modello è una variabile continua tra 0 e 1 che può essere interpretata come probabilità di default di un singolo interlocutore in un arco temporale prestabilito, dati determinati input. Utilizzando questo modello, si potrà rispondere a domande tipo:

1. Come può essere utilizzato questo modello per ottimizzare le decisioni di fornire credito?
2. Qual'è il tasso aggregato di insolvenza equali sono le sue conseguenze sull'andamento del ciclo economico?

Una delle misure naturali di capacità predittiva dell'approccio descritto consiste nel confrontare gli score degli individui che hanno fatto default sui loro debiti con quelli di coloro che invece hanno onorato le loro obbligazioni. Differenze significative tra i diversi score certificherebbero che le previsioni hanno un genuino potere di discriminare tra i due gruppi di debitori.

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE

OR2.1 Creazione data base integrato

Obiettivo: L'attività ostituisce uno step preliminare fondamentale per trasferire in un unico database gli input e gli output che saranno poi utilizzati nella fase della modellistica.

Organizzazione: Particolare attenzione verrà data all'analisi dei casi in cui ci sono dati discordanti per lo stesso interlocutore. Saranno individuate procedure automatiche per risolvere eventuali conflitti.

Tempistica: Questo database sarà successivamente integrato dalle risultanze dell'analisi testuale delle registrazioni telefoniche.

mese di inizio	1
mese di fine	30
durata (in mesi) dell'attività	30

Consegne:

1. Versione 1 del data base integrato: $T0 + 6$ mesi;
2. Versione 2: $T0 + 18$ mesi;
3. Versione finale: $T0 + 30$ mesi.

OR2.2 Principali determinanti del default

Obiettivo: La relazione che definisce la cosiddetta perdita attesa (Expected Loss, EL) relativa ad un finanziamento nei confronti di un determinato interlocutore è la seguente:

$$EL = PD \times LGD \quad (7.1)$$

dove: Probability of default (PD) è la probabilità che l'interlocutore non onori il suo debito e diventi insolvente. L'ultimo elemento è la loss given default (LGD) e si riferisce ad una quantificazione della perdita connessa a ciascun finanziamento, condizionata al fatto che il debitore diventi insolvente.

Visto che la variabile output di ciascuna pratica è binaria, nel senso che può assumere soltanto due valori (default, non default), questo rappresenta un problema di classificazione.

Organizzazione: Utilizzeremo classification and regression trees (CART) per costruire le nostre misure predittive. CART è una procedura statistica dove una variabile output dipende da un set di input attraverso una sequenza recursiva di relazioni semplici binarie (una sorta di albero decisionale). L'intera collezione di tutte le relazioni recursive riesce a partizionare lo spazio multi-dimensionale degli input in regioni distinte dove gli output sono assunte essere costanti.

I parametri del modello sono scelti in base ad un processo di minimizzazione della distanza tra i valori output ed le misure fittate dal modello utilizzando una appropriata misura di distanza (tipicamente l'errore quadratico medio). Da qui si possono costruire le probabilità di default future.

La crescente popolarità dell'approccio CART sta nel fatto che questo supera tutte le limitazioni dei modelli tradizionali logit e probit dove la relazione tra variabile output e gli input è imposta essere lineare su tutto lo spazio delle variabili input. Al contrario, CART è capace di svelare relazioni non lineari tra le variabili input, cosa che aumenta drammaticamente il tipo di relazioni che possono essere descritte ed il numero di input che può essere utilizzato. Inoltre, i modelli CART producono regole decisionali chiare ed intuitive la cui logica è catturata dall'albero decisionale.

Tempistica:

mese di inizio	7
mese di fine	24
durata (in mesi) dell'attività	18

Consegne:

1. Prima serie di determinanti del default: $T0 + 12\text{mesi}$
2. Serie finale: $T0 + 24\text{mesi}$.

OR2.3 Principali Determinanti della LGD

Obiettivo: La LGD è di gran lunga la componente più difficile da determinare, soprattutto perché sono molto più scarse le "osservazioni" statistiche disponibili. Si tratta di un problema analiticamente complesso, in quanto altamente non lineare, multidimensionale e in quanto il database è di tipo non strutturato. Pertanto le analisi tradizionali derivanti dalla statistica classica non sono adatte ad affrontare un problema di questo tipo. I modelli lineari non sono in grado di catturare le dinamiche non lineari presenti nei dati. Ma anche i modelli non parametrici non possono essere utilizzati perché il processo è troppo multidimensionale (ed il database non strutturato) per pervenire a delle stime statisticamente attendibili.

L'idea è di sfruttare le recenti tecniche di analisi di classificazione utilizzando strumenti di machine learning e tecniche che penalizzano (in senso statistico) la multidimensionalità del modello e che quindi tendono a scegliere il modello più parsimonioso per descrivere il processo che genera i dati. Machine learning è particolarmente adatta ad analizzare database di questo tipo (non strutturati, multidimensionali e molto grandi) per cui l'aspettativa è di migliorare notevolmente la capacità predittiva di profilazione del rischio e di rendimento atteso da ogni finanziamento.

Le operazioni descritte saranno alla base di un nuovo software applicativo di calcolo della LGD, che avrà 2 funzioni principali:

1. Calcolare la aspettativa non condizionata della LGD, che non tiene conto della probabilità di insolvenza di ciascun interlocutore;
2. Calcolare l'aspettativa condizionata della LGD, che internalizza la probabilità di default di ciascun interlocutore, calcolata in maniera robusta, quindi non necessariamente imponendo la fattorizzazione indicata per ragioni di semplicità espositiva nella formula (7.1);
3. Calcolare la relazione rischio/rendimento: il modello di calcolo della LGD unito ai modelli di credit scoring per il calcolo della PD, consentirà di realizzare un compiuto modello di calcolo per la relazione rischio/rendimento che è alla base della misurazione del più corretto livello di tasso di interesse in relazione al rischio di credito assunto dal lender.

Ci possono essere determinanti che hanno effetto sia sulla PD che sulla LGD, quindi se questo è il caso la fattorizzazione non è corretta. E pertanto richiede l'utilizzo di tecniche non parametriche che si adattino al tipo di database indicato. Il ricorso a tecniche di big data e artificial intelligence è, anche in questo caso, quanto mai opportuno.

Al momento (2018) FIRE ha iniziato un progetto di ricerca finalizzato ad individuare quali variabili - che riguardano tanto caratteristiche relative alla pratica stessa, quanto invece relative all'interlocutore, quanto infine allo stato generale della situazione economica del Paese - influenzano una misura di redditività della soddisfazione della pratica stessa, che FIRE misura col rapporto tra somme realizzate e somme affidate. Per costruzione, tale rapporto varia tra 0 e 1, raggiungendo 1 quando tutto il credito viene incassato. FIRE ha selezionato un campione all'interno dell'universo delle pratiche gestite, e ha calcolato misure di associazione univariate tra ogni singola variabile esplicativa ed il rapporto tra realizzato ed affidato.

Si tratta di un primo step, che sarà migliorato ed ottimizzato secondo le linee seguenti. In primo luogo, la misura di redditività relativa ad ogni singola pratica sarà modificata in maniera tale da includere anche misure relative ai costi di gestione della singola pratica. In maniera tale da arrivare ad una misura che sia indicativa della profittabilità di ogni singola pratica. In secondo luogo, la complessità del database invece di costituire un ostacolo all'analisi sarà sfruttata come un vantaggio. Si procederà ad una analisi multivariata delle varie determinanti del successo di ogni singola pratica, attraverso tecniche di machine learning. Tali tecniche sono particolarmente indicate quando il database a disposizione è sparso e non strutturato, com'è il caso in questione. Le tecniche di machine learning consentono di gestire database di grande dimensione (big data) e permettono di individuare forme di associazione non lineare tra le singole variabili, cosa che non è possibile utilizzando i classici modelli multivariati.

Organizzazione: Le tecniche utilizzate spazieranno CART a modelli più complessi di deep learning. L'analisi partirà con lo sfruttamento del database esistente. Nella seconda fase del progetto, non appena sarà completata la fase di analisi semantica delle registrazioni telefoniche compiuta da Expert Systems, lo studio includerà anche questo importante input. Un aspetto innovativo rispetto alla letteratura esistente della attività di ricerca in questa fase del progetto è che il dizionario attraverso il quale interpretare i contenuti testuali delle registrazioni telefoniche non sarà imposto dall'esterno (per evitare distorsioni interpretative dei singoli termini o paragrafi), ma bensì sarà estratto dai dati stessi, visto che si è in possesso della variabile che esprime il grado di profittabilità di ogni singola pratica. Il risultato atteso da questo esercizio è duplice:

1. Separare l'effetto delle caratteristiche della singola pratica o dell'interlocutore dal modo in cui l'attività di riscossione del credito viene esercitata (cosa che è possibile fare grazie all'analisi semantica del contenuto delle registrazioni telefoniche) e dallo stato della pratica.
2. Individuare esempi di best-practice su come gestire ogni singola pratica a seconda dello stato della pratica e delle caratteristiche dell'interlocutore. Questi esempi saranno trascritti in protocolli che ciascun singolo operatore potrà seguire nelle varie fasi di gestione della pratica.

MIRC.0

Tempistica:

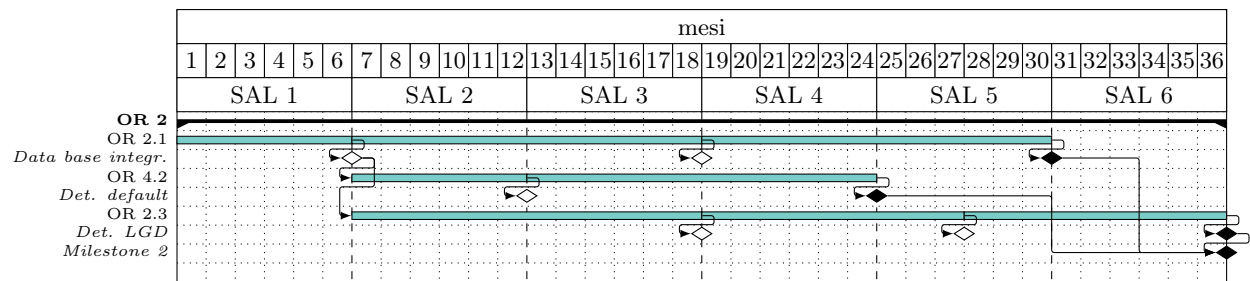
mese di inizio	7
mese di fine	36
durata (in mesi) dell'attività	30

Consegne:

- 1. Prima serie di determinanti della LGD: T0 + 18 mesi;
- 2. Seconda serie: T0 + 27 mesi;
- 3. Serie finale: T0 + 36 mesi.

TEMPI DI REALIZZAZIONE

Durata (in mesi): 36



7.3. OR3 – Applicazione, testing e fine tuning degli algoritmi

DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO

Titolo dell'OR: Applicazione, testing e fine tuning degli algoritmi in un dimostratore

Soggetto proponente: UNIME

Tipologia (RI/SS): SS

Tempistica dell'OR:	mese di inizio	1
	mese di fine	36
	<hr/> durata (in mesi) dell'OR	<hr/> 36

Luoghi di svolgimento: Sede UNIFE

Obiettivo: Sviluppare dimostratori ed effettuarne una convalida e verifica sperimentale.

Parametri che permettono di misurare il raggiungimento dell'obiettivo: In tale attività si vuole sviluppare, validare e testare un dimostratore del sistema di classificazione automatica, e un dimostratore del sistema di ML non supervisionato. Nel seguito, si denomina *miner* quello basato su tecniche di ML non supervisionato e *classifier* quello basato su tecniche di ML supervisionato. Si svolgerà sia la validazione (con supporto di esperti di Fire) della conoscenza estratta dal *miner*, sia una valutazione della bontà della classificazione del dimostratore *classifier*. Si attuerà una verifica sperimentale, da applicare su casi identificati nella prima fase di OR3. L'applicazione sarà valutata in termini di significatività dei cluster e delle regole associative scoperte con tecniche di ML non supervisionato. Per il dimostratore *classifier* (basato su alberi decisionali e regole di classificazione), la conoscenza estratta sarà validata dagli esperti (per poterla anche integrare poi nel sistema di supporto alle decisioni a regole), e saranno misurate accuratezza, percision e recall sui casi identificati.

Tipo di intervento:

1. Realizzazione di un dimostratore *miner* per scoperta (mining) dai dati (basato clustering o regole associative)
2. Realizzazione di un dimostratore *classifier* per classificazione dei dati (basato su alberi decisionali o regole classificatorie)
3. Valutazione dei dimostratori prototipali (validazione e prestazioni)

Principali problematiche da affrontare:

1. Identificazione dei casi da utilizzare per validazione e valutazione delle prestazioni
2. Identificazione delle implementazioni più idonee degli algoritmi identificati in OR1
3. Messa in opera di prototipi dimostratori facilmente aggiornabili e manutenibili;
4. Applicazione dei dimostratori ai casi individuati (validazione e prestazioni), e fine tuning.

Tempi e modalità di realizzazione dell'intervento:

1. Scelta delle implementazioni degli algoritmi di ML, basandosi su tecnologie open-source o di dominio del gruppo di ricerca;
2. Sviluppo sperimentale di un dimostratore *miner*;
3. Sviluppo sperimentale di un dimostratore *classifier*.

Test da effettuare per misurare i parametri e tempi di realizzazione: Secondo la lista delle attività riportata

Note descrittive generali: Si svolgeranno due tipi di sperimentazioni con i dimostratori sviluppati:

1. la prima sperimentazione ha l'obiettivo di validare: (i) la conoscenza estratta dal *miner* e quindi valutare - con l'aiuto di esperti del settore - la significatività della conoscenza estratta automaticamente dai dati (con clustering o scoperta di regole associative, o apprendimento di regole classificatorie e alberi decisionali); (ii) la conoscenza alla base del *classifier* appreso (rappresentata come un albero decisionale o con regole). Si tratta quindi di una valutazione umana, qualitativa;
2. la seconda sperimentazione ha l'obiettivo di validare la bontà del dimostratore classificatore sviluppato (basato su alberi decisionali o regole dclassificatorie) su casi non visti in OR1, e quindi verificare ulteriormente la capacità predittiva su esempi che non fanno parte del test set considerato in OR1, e che emergano come significativi nel corso di OR3 (per almeno un 10% ulteriore di dati non visti e non facenti parte del test set), misurando l'accuratezza ottenuta. Obiettivo è avere su questi nuovi dati almeno accuratezza > 80 .

Le attività dell'OR si collocano nell'ambito dello sviluppo sperimentale. Si adotteranno - ove disponibili - realizzazioni open-source di algoritmi di ML non supervisionato e supervisionato, per lo la messa in opera dei dimostratori. Ad esempio, il sistema Weka, sviluppato in Java, raccoglie una suite di algoritmi di ML supervisionato e non, realizzati in tale linguaggio, e può costituire un'utile base allo sviluppo dei dimostratori.

Il gruppo di UNIFE ha inoltre sviluppato algoritmi di apprendimento di regole probabilistiche la cui implementazione è disponibile per la realizzazione di un dimostratore classificatorio, che integri anche una misura in termini di probabilità della classificazione.

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE

OR3.1 Definizione degli esempi per convalida e verifica sperimentale dei dimostratori

Obiettivo: Identificare una classe sufficientemente ampia di esempi, adottando criteri di copertura opportuni

Organizzazione: Tale attività sarà fondamentale, per individuare e raccogliere ove non già disponibili, casi di test particolarmente significativi al fine di coprire tutte le tipologie di pratiche e/o di interlocutori per il dimostratore classificatore. Sarà utile che gli esempi comprendano anche registrazione delle conversazioni (vocale o trascritta delle conversazioni), in modo da poterle analizzare negli altri OR (OR4 e OR5).

Per il dimostratore estrattore di conoscenza, e basato su algoritmi di ML non supervisionato, gli esempi da identificare in OR3.1 saranno espressi in termini di "informazioni" o conoscenza che gli esperti desidererebbero scoprire dai dati (ovvero proprietà, quali ricorrenze, correlazioni, caratteristiche di gruppi omogenei, che si desidera verificare se estraibili dai dati).

Pertanto, con il termine "esempio" in OR3.1 si coprono entrambe le cose (nuovi casi da classificare, e proprietà).

Tempistica:

mese di inizio	1
mese di fine	12
durata (in mesi) dell'attività	12

Consegne:

1. Identificazione "esempi" per convalida e verifica sperimentale dei dimostratori: T0 + 12 mesi

OR3.2 Sviluppo e applicazione dei dimostratori

Obiettivo: Realizzare dimostratori e validarli sperimentalmente

Organizzazione: In tale attività si realizzeranno due dimostratori, uno basato su tecniche di ML non supervisionato (dimostratore *miner*) e uno su tecniche di ML supervisionato (dimostratore *classifier*). I due dimostratori saranno basati sugli algoritmi/strumenti identificati in OR1 e saranno applicati agli “esempi” di OR3.1.

L’esito sarà analizzato anche con gli esperti di dominio. A valle di tale attività si effettuerà una retroazione, e un fine tuning dei sistemi e degli algoritmi alla loro base.

Tempistica:

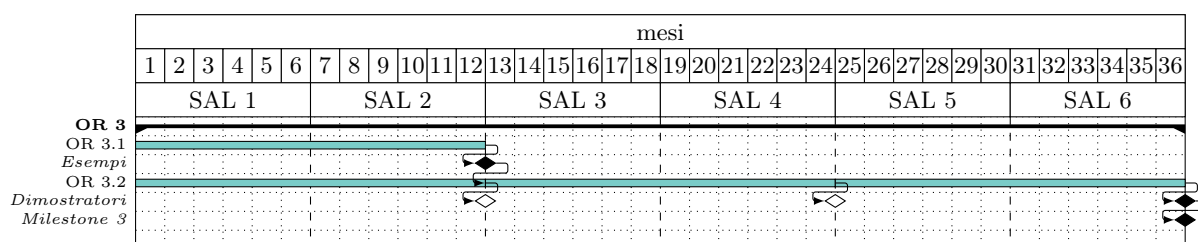
mese di inizio	1
mese di fine	36
durata (in mesi) dell’attività	36

Consegne:

1. Definizione GUI dei dimostratori: $T_0 + 12$ mesi ;
2. Integrazione algoritmi ML nei dimostratori e prima validazione: $T_0 + 24$ mesi
3. Retroazione, e seconda validazione sperimentale utilizzando anche informazioni estratte dalle conversazioni relative agli esempi da OR4 e OR5: $T_0 + 36$ mesi

TEMPI DI REALIZZAZIONE

Durata (in mesi): 36



7.4. OR4 – Creazione base dati con informazioni automaticamente estratte da contenuti non strutturati.

DESCRIZIONE DELL’OBIETTIVO REALIZZATIVO

Titolo dell’OR: Creazione base dati con informazioni automaticamente estratte da contenuti non strutturati.

Soggetto proponente: Expert System

Tipologia (RI/SS): RI

Tempistica dell’OR:	mese di inizio	1
	mese di fine	34
	durata (in mesi) dell’OR	34

Luoghi di svolgimento: Sede Expert System

Obiettivo: Sviluppare soluzioni di Intelligenza Artificiale per estrarre conoscenza dalla moltitudine di contenuti non strutturati (come file audio delle conversazioni tra operatori e debitori o altri documenti testuali) al fine di creare una base dati strutturata a supporto della profilazione degli stakeholders coinvolti nel processo di recupero del credito e della automazione dei suggerimenti personalizzati e/o a supporto dei processi della fase di contatto tra operatore e interlocutore.

Parametri che permettono di misurare il raggiungimento dell'obiettivo:

1. **Analisi automatica di varie tipologie di fonti (es. file audio, report, verbali, txt, pdf, doc, ...);**
2. Categorizzazione ed estrazione automatica di informazioni rilevanti, individuate secondo le indicazioni degli esperti di settore;
3. Scalabilità dei processi per una piena integrazione del sistema nella piattaforma operativa Fire al fine di contribuire al suggerimento delle azioni più efficaci nella futura attività degli operatori coinvolti.

Tipo di intervento:

1. **Identificazione e creazione di Corpora di file audio e documenti significativi e rappresentativi del dominio di conoscenza da mappare;**
2. **Implementazione dei motori di analisi vocale e testuale;**
3. **Arricchimento** dei contenuti con metadati qualitativi e quantitativi connessi all'analisi linguistica (es. principali concetti, tipo di linguaggio usato, mood della conversazione, ...)
4. Esecuzione dei test e raccolta feedback per il fine tuning del sistema;
5. Progettazione e sviluppo di un dimostratore del sistema di analisi semantica.

Principali problematiche da affrontare:

1. **Comprensione e trascrizione automatica di file audio con diversi parlatori, possibile utilizzo di dialetti e qualità del segnale ridotta;**
2. **Comprensione ed estrazione automatica di informazioni da file di testo generati dal riconoscimento di file audio (ASR)**
3. Comprensione ed estrazione automatica da diverse tipologie di documenti (es. pdf, tabelle, xls,...) di informazioni espresse nei diversi modi possibili, tipici del linguaggio di tutti i giorni, che meglio rappresentino sinteticamente la conoscenza del dominio.
4. Identificazione ed estrazione delle categorie ed entità utili per una precisa descrizione del dominio di riferimento (metadati semantici), compresa l'eventuale definizione di una Tassonomia di dominio.

Tempi e modalità di realizzazione dell'intervento: Secondo la lista delle attività di seguito riportata.

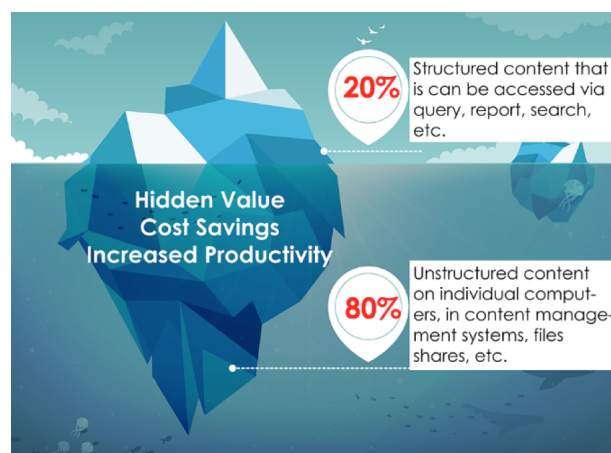
Test da effettuare per misurare i parametri e tempi di realizzazione: In ambito del NLP, l'accuratezza dei risultati si misura principalmente con i valori di precision e recall.

Il primo indica la correttezza dell'analisi, vale a dire se il processo automatico ha correttamente disambiguato uno specifico termine o concetto. Il secondo indica l'eventuale presenza di falsi positivi o falsi negativi. Nel corso del progetto si procederà con:

1. **misurazione precisione del riconoscimento vocale dei file audio, da attestarsi su un valore di Error Rate su canale telefonico pari al 20%.**

2. misurazione precision e recall dell'analisi linguistica, con obiettivo di precision > 70% e recall > 65%.
3. Verifica della capacità predittiva degli esiti positivi o negativi di processi "real time" effettuata sulla base del corpus storico (serie storica di dati taggati).

Note descrittive generali: Oltre l'80% della conoscenza è contenuta in files non strutturati, di tipo testuale. La quantità di dati è talmente elevata che lo sforzo per analizzare / comprendere manualmente tali contenuti è insostenibile, se non in modo mirato e sporadico. Solo l'adozione di sofisticate tecniche di Intelligenza Artificiale permettono di affrontarne l'analisi in modo continuativo e strutturato.



Le attività di questo OR si collocano nella cosiddetta Robotic Process Automation, cioè uno dei 3 principali ambiti applicativi della AI, insieme al Cognitive Insight e al Cognitive Engagement. L'approccio linguistico profondo non è generalmente utilizzato in quest'ambito perché prevede un intenso e costoso lavoro manuale di creazione della base dati linguistica (knowledge graph) che l'approccio statistico non richiede (almeno in teoria) se è disponibile una grande quantità di dati (corpus).

L'approccio linguistico permette tuttavia di ottenere migliori performance in termini di Precision rispetto al Machine Learning ed in un ambito di RPA con intensa interazione con gli utenti finali, l'esatta comprensione dei contenuti è un prerequisito assoluto.

Grazie al proprio knowledge graph, già estremamente ricco di concetti e conoscenza, l'analisi Cogito non necessita di un grande corpus taggato per addestrare il sistema. Ciò semplifica molto lo sviluppo, velocizza i tempi dei rilasci e limita i costi (anche di gestione futura).

Expert System ha investito oltre 150 anni/uomo per creare il Knowledge Graph della lingua italiana, chiamato Sensigrafo. Esso è uno strumento unico, sviluppato con tecniche divenute oggetto di brevetto negli Stati Uniti, che già ha in sé un'estesa mappa della conoscenza e permette "out of the box" di comprendere ogni tipo di contenuto. Non c'è bisogno quindi di grandi moli di dati per portare la sua conoscenza da 0 ad un livello operativo, ma sono invece sufficienti pochi documenti "mirati", per arricchire quanto già disponibile.

Ciò rappresenta una peculiarità di grande valore dell'approccio qui proposto, perché di fatto rende più semplice lo sviluppo e l'applicazione delle soluzioni evolute di AI attraverso un approccio iterativo (sviluppo > rilascio > misurazione) AGILE.

Funzionalmente, L'approccio linguistico profondo sarà collocato in una pipeline qui schematizzata in Figura 7.1.

Grazie all'analisi linguistica e comprensione dei testi, sarà possibile popolare un archivio (rappresentato in colore rosso in Figura 7.1) composto da dati strutturati (metadati semantici) che "distillano" la conoscenza contenuta in milioni di files e contenuti testuali o vocali e dati quantitativi ad essi collegati, di cui si riporta un esempio non esaustivo della possibili tipologie nella tabella 7.1.

L'analisi semantica profonda, comprende varie attività di elaborazione del testo, propedeutiche all'estrazione degli indicatori o delle informazioni rilevanti (o pepite informative). Un' analisi di base è svolta dallo strato della tecnologia linguistica Cogito in grado di effettuare una analisi del testo dal punto di

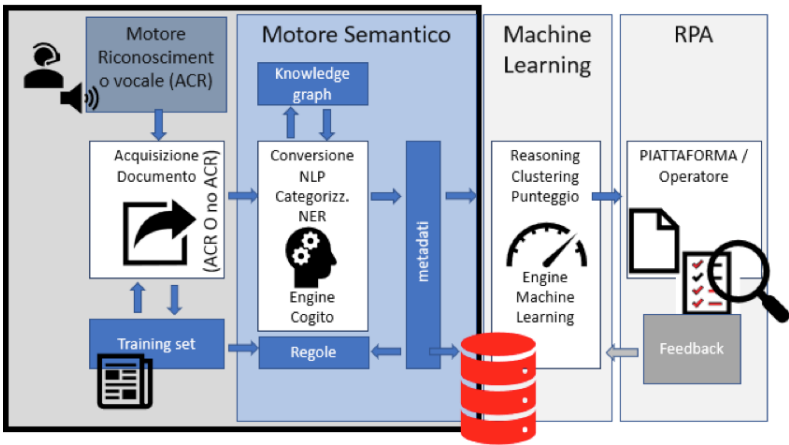


Figura 7.1: Pipeline OR

	TIPO DATI	ESEMPLI ITEMS
A	Dati Semantici rappresentativi dei contenuti testuali	Main sentences Main Concepts Categoria associata a Tassonomia di dominio Entità SAO (analisi Soggetto Azione Oggetto) Relazioni tra entità ...
B	Dati Semantici rappresentativi del profilo dell'autore	Ricchezza linguaggio Utilizzo slang / termini scurrili Mood (positivo o negativo) del testo ...
C	Dati Quantitativi	Numero parole Orario registrazione (in caso di file audio) ID Pratica (se esistente) Ruolo parlatori (in caso di file audio)

Tabella 7.1: Esempio della possibili tipologie di dati.

vista grammaticale, sintattico e semantico. L'integrazione della suddetta analisi con il Sensigrafo, permette di attivare l'analisi semantica profonda, vale a dire il processo di disambiguazione ⁽¹⁾. A questi, si aggiunge lo strato di regole linguistiche alla base dell'analisi custom di Cogito (Rules Based), derivato dalla produzione semi-automatica delle regole linguistiche di match che hanno come operandi gli elementi di output dell'analisi di base. Le regole infatti sono basate sull'output dell'analisi NLP del testo, non sul testo originario, quindi sono ad un livello di astrazione più alto perché terranno conto delle informazioni grammaticali, sintattiche e semantiche. In sostanza le regole non saranno scritte per fare match direttamente sul testo bensì sugli elementi di output delle analisi del testo stesso. Questo permetterà, ad esempio, di scrivere regole che scatteranno in presenza di espressioni di determinati concetti (non keyword) oppure quando un certo termine risulterà essere, rispetto all'analisi sintattica, il soggetto di una proposizione e non l'oggetto, ecc. Il motore analitico sarà sviluppato da Knowledge Engineer per rilevare automaticamente gli indicatori presenti nei testi, estrarli e restituirli come info di output, per popolare il data base della conoscenza di dominio. Si tratta di valori normalizzati in tecnica ETL (Extraction, Transformation and Loading) e il procedimento di estrazione è quello della Named Entity Recognition (NER) dove tutte le parti di testo che possono dare indicazioni di concetti/procedure (singoli termini, forme

¹Disambiguare una parola, significa associarne il corretto significato in quel contesto. Ad esempio nell'espressione "ti rifilo un calcio", calcio significa "pedata" e non "football" o "calcio della pistola", ecc..

base di sostantivi/verbi/aggettivi, collocazioni, coppie nome-valore o range di valori, concetti, espressioni tipiche o intere proposizioni...) saranno individuate mediante regole di matching redatte in modalità semi-automatica (supervised).

In output verrà emesso, insieme al brano di testo su cui scatta la regola, un valore rappresentativo normalizzato. Ad esempio, dalla frase: *Nella giornata di domani volevo mandare una raccomandata a Pippo srl per far sì che Luca rimuova il tutto, dato che il servizio è stato cessato in data 2/09/18*

Verranno estratti i dati in formato normalizzato:

DATE = domani (= today+1); 02/09/2018

AZIONE = invio raccomandata

CATEGORIA = cessazione di servizio

ORGANIZATION: Pippo srl

PEOPLE: Luca

MOOD: neutro

I valori normalizzati potranno essere resi conformi ad una convenzione o standard indicato dagli stakeholders con cui si collabora. Il valore di default adottato per la categorizzazione sarà la Tassonomia Finance di Cogito. L'estrazione potrà essere anche aggregata, quando è possibile associare un evento/concetto ad un altro. Ad esempio, sempre partendo dal testo sopra:

ESTRAZIONE 1 = (domani) ! invio raccomandata

ESTRAZIONE 2 = cessazione servizio ! 02/09/2018

insieme alle coordinate del brano di testo (la posizione), utile per l'eventuale evidenziazione a video o in una versione annotata del documento.

Questo dato è più significativo dei singoli dati disaggregati perché questi possono indicare una generica co-occorrenza nel testo mentre il dato aggregato è indicativo di una correlazione, che risulta potenzialmente più promettente per l'individuazione dei pattern nella successiva fase di analisi ML.

Inoltre, una volta strutturato in questo modo, tale dato può essere facilmente messo in relazione con altre informazioni strutturate presenti in Azienda, al fine ad esempio di verificare l'autenticità di taluni valori. Assumendo che il file di testo utilizzato come esempio, sia riconducibile ad un codice identificativo univoco della pratica, sarà possibile verificare se la data di cessazione indicata dal debitore coincide o meno con quella presente nei propri archivi.

Zoomando sugli aspetti di originalità dell'analisi dei contenuti da analizzare, si evidenzia la complessità dell'analisi di un testo generato dall'analisi vocale (ASR), che presenta peculiarità nel lessico ma anche nella sintassi e nella punteggiatura (assente), la cui comprensione necessita lo studio ed attivazione di specifiche strategie di analisi del linguaggio. Oltre a ciò, i testi riportano di frequente espressioni dialettali o "rumore" generato dal parlato spontaneo (esempio colpi di tosse, mugugni) dall'esito non scontato i fase di riconoscimento.

A titolo di esempio, si osservi la seguente tabella 7.2, è evidente che l'analisi del testo riportato in colonna A o dello stesso testo rappresentato in colonna B, implica strategie e complessità assai diverse. Premesso che la colonna B rappresenta un tipico testo "piatto" di output di un ASR standard, nel corso del progetto si studieranno metodologie per ricondurlo automaticamente il più possibile al testo di colonna A (con distinzione del parlatore, punteggiatura,...)

Infine non scontato sarà l'esito dell'utilizzo di dati generati dal Machine Learning e dall'analisi semantica, per arricchire ed organizzare la conoscenza necessaria all'elaborazione di efficaci cluster e suggerimenti operativi. Tale approccio (ML+NLP) rappresenta una sfida tra le più promettenti nell'ambito della Ricerca in ambito "content analysis", e si colloca per questo nel filone della ricerca più attuale e promettente, in corso in tutto il mondo. Da questo punto di vista è importante rilevare come la disponibilità di un Corpus molto ampio di contenuti di cui si conosce già l'esito (set storico) rappresenta un grande valore per il progetto e fornisce estrema concretezza all'esplorazione delle variabili (tipiche o totalmente inusuali per il dominio), alla ricerca delle combinazioni più efficaci (e non scontate) di rappresentazione della conoscenza in funzione predittiva.

COLONNA A	COLONNA B
OP 1: Ho capito. Va bene, allora chiami e dica loro di togliere in giornata la pratica a noi come Fire, va bene?	Ho capito va bene allora chiami e dica loro di togliere in giornata la pratica a noi come Fire va bene sì sì di fare quel che va fatto per chiudere
USER: Sì sì, di fare quel che va fatto per chiudere il tutto...	il tutto va bene grazie buona serata salve.
OP 1: Va bene. Grazie buona serata, salve.	

Tabella 7.2: Esempio di testo con doppia interpretazione.

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE

OR4.1 Creazione Corpus file audio e documenti taggati

Obiettivo: Definire e implementare corpora annotati

Organizzazione: Tale attività sarà ridotta al minimo, tuttavia è di grande importanza allestire da subito un Corpus di documenti e file audio da utilizzare per lo sviluppo ed il fine tuning del sistema, come training set e nelle fasi di benchmarking e controllo nel rilascio delle versioni evolutive del software. Particolare attenzione verrà data al corpus di file audio. Fire registra circa 4 milioni di conversazioni operatore – debitore all'anno. Da questo insieme verranno creati 2 corpus per lo sviluppo e test delle performance del riconoscimento vocale e del text mining.

Una parte dei file verranno quindi trascritti manualmente al fine di creare il set di test su cui far girare il motore per misurare le performance in termini di precisione ed accuratezza.

Le principali caratteristiche file audio sono:

- Segnale Mono;
- Due parlatori;
- Linguaggio spontaneo > possibili sovrapposizioni;
- Lingua Italiano / dialetti;
- Disponibilità Metadati (es orario file, ID pratica...).

Se i testi trascritti da ASR appaiono i più promettenti, tuttavia anche una parte di altre tipologie di documenti verranno taggate manualmente al fine di creare il set di test su cui far girare il motore NLP per misurare le performance in termini di precisione e recall e capacità predittiva.

Tempistica: Si prevede un costante arricchimento del suddetto Corpus nel corso del primo anno di progetto, tuttavia nei primi 6 mesi ci si attiverà in modo particolarmente intenso su questi aspetti.

mese di inizio	1
mese di fine	30
durata (in mesi) dell'attività	30

Consegne:

1. Versione 1 del corpus files Audio e del corpus documenti: $T_0 + 6$ mesi;
2. Versione 2: $T_0 + 12$ mesi;
3. Versione finale: $T_0 + 30$ mesi.

OR4.2 Riconoscimento file audio (ASR)

Obiettivo: Riconoscimento e trascrizione automatica di registrazioni di colloqui operatore-debitore.

Organizzazione: Premesso che un motore ASR gestisce i file contenenti audio in formato non compresso wav, PCM mono e stereo 16/24 bits, con frequenza di almeno 8 o 16 KHz e contenenti formati compressi MP2, MP3, MP4, MPG, con bitrate preferenziale di 256 kbps, l'output viene generalmente riportato in file XML, che contiene il testo della telefonata suddiviso in frasi e (se presenti) suddiviso per canale.

Il componente effettua la trascrizione automatica utilizzando un Modello del Linguaggio (Language Model) e un Modello Acustico (Acoustic Model) che possono essere aggiornati o rigenerati su materiale fornito dal cliente. A seguito della raccolta di un quantitativo utile di dati (registrazioni) forniti da Fire, sarà quindi possibile effettuare una personalizzazione del modello del linguaggio migliorando l'accuratezza del sistema sul lessico e sul linguaggio tipico del dominio in oggetto. Il testo inoltre potrà essere accompagnato dai relativi riferimenti temporali e dalle informazioni sulla confidence e sulla pronuncia. I tempi necessari per la trascrizione automatica di telefonate sono nell'ordine di un rapporto 1 : 1. Il delay della trascrizione dipende dal numero di istanze di trascrizione utilizzate, dalla quantità di audio inviato e dalla frequenza dell'invio di audio da parte del committente. Durante il progetto si effettueranno delle analisi offline, non si prevede l'integrazione del motore in modalità realtime, ma il sistema sarà scalabile e potrà essere opportunamente dimensionato per trascrivere il 100% del volume delle telefonate (oltre 4 milioni/anno) anche in questa modalità.

Le principali peculiarità del motore di trascrizione sono:

- Speaker Independence;
- Natural Language;
- Continuous Speech Recognition;
- Vocabolari molto grandi (300k+ parole);
- Speech Recognition robusto.

Il modello acustico può essere a 8kHz.

Per tale attività ci si prevede l'adozione di un motore ASR disponibile sul mercato, individuato tra quelli più adatti al tipo di riconoscimento qui necessario.

Tempistica:

mese di inizio	4
mese di fine	24
durata (in mesi) dell'attività	21

Consegne:

1. Primo Corpus file trascritti (circa 10%): T0 + 6 mesi
2. Secondo Corpus file trascritti (circa 50%): T0 + 12 mesi
3. Corpus di test completo (100%): T0 + 24 mesi.

OR4.3 Analisi linguistica profonda dei testi

Obiettivo: Analisi semantica di contenuti non strutturati (prodotti o non prodotti da ASR).

Organizzazione: Con NLP (Natural Language Processing) si indica il trattamento automatico di informazioni testuali scritte in linguaggio naturale, di tutti i giorni, che a sua volta si compone di più fasi: Le diverse fasi dell’NLP, categorizzazione dei contenuti, estrazione delle entità, mood extraction, writeprint, sono operazioni che richiedono un’elevata efficienza dell’algoritmo, per garantire un ottimo livello prestazionale che può essere quantificato mediante i parametri “precision” e “recall”.

La disponibilità di tutte le suddette informazioni renderà estremamente più ricca ed efficace la possibilità di clusterizzare le tipologie di interazione o gli stakeholder coinvolti, in insiemi molto precisi ed efficaci. Tanto più che attualmente vengono definite solo macro-categorie molto “ampie”, basate sulla grezza distinzione tra cliente business/consumer o sulla scadenza del debito maggiore o minore di 30 gg.

Grazie a Cogito si creeranno nuovi cluster molto più piccoli / personalizzati che permetteranno di suggerire agli operatori delle scelte più precise (anche se standardizzate, quindi non soggettive). L’accento sarà posta sui contenuti presenti nelle trascrizioni dei file audio delle conversazioni tra operatori e debitori (diversi milioni / anno), tuttavia vi sono una serie di altri documenti collegati ad una singola pratica (vedi figura 7.2) che possono contenere ulteriori interessanti informazioni per arricchire la rappresentazione della pratica lato debitore, ma anche lato operatore, allo scopo di individuare quegli elementi che possono far emergere atteggiamenti inopportuni o criticità, come ad esempio, la ricorrente assenza di un consulente a cui un operatore call center tenta di girare la chiamata o l’utilizzo di un linguaggio inappropriato.

Infine, la “pratica standard”, composta dai Report, in cui vengono riportate dal PM in forma schematica le fasi di contatto e gli esiti degli stessi (in giallo sono evidenziati i contatti per i quali sono disponibili le registrazioni audio), data set presenti nel DB / CRM aziendale, ecc. potrebbe essere arricchita con informazioni esterne, automaticamente ricercate sul WEB, attraverso forme di Open Source Intelligence.

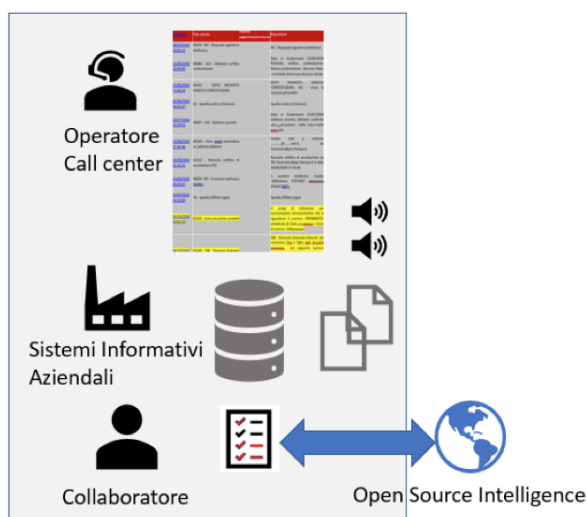


Figura 7.2: Ricchezza di contenuti non strutturati di una “pratica di recupero crediti”

Tempistica:

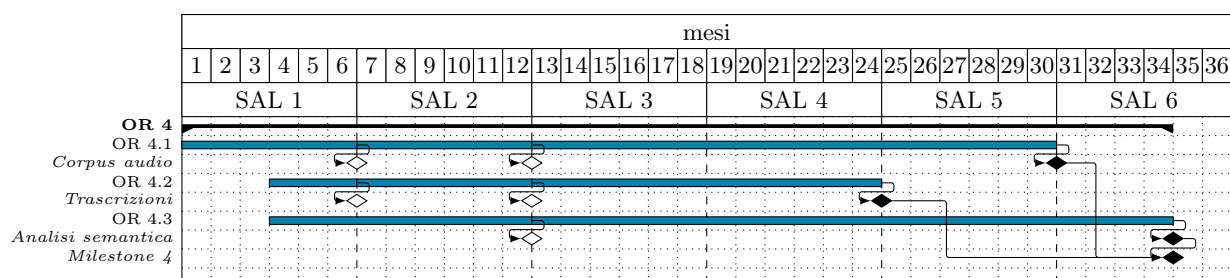
mese di inizio	4
mese di fine	34
durata (in mesi) dell’attività	31

Consegne:

1. Versione 0: T0 + 12 mesi;
2. Versione 1: T0 + 24 mesi;
3. Versione finale: T0 + 34 mesi.

TEMPI DI REALIZZAZIONE

Durata (in mesi): 34



7.5. OR5 – Estrazione dal testo di informazioni per la profilazione puntuale del soggetto analizzato

DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO

Titolo dell'OR: Estrazione dal testo di informazioni per la profilazione puntuale del soggetto analizzato

Soggetto proponente: Expert System

Tipologia (RI/SS): SS

Tempistica dell'OR:	mese di inizio	1
	mese di fine	18
	durata (in mesi) dell'OR	18

Luoghi di svolgimento: Sede Expert System

Obiettivo: Analisi semantica integrata con profilo multidimensionale degli interlocutori.

Parametri che permettono di misurare il raggiungimento dell'obiettivo:

1. Capacità di individuare cluster ben riconoscibili / distinti;
2. Capacità di estrarre informazioni significative generate dall'analisi Semantica.

Tipo di intervento:

1. Identificazione ed approfondimento delle peculiarità dei vari ambiti applicati / tipo di problematica;
2. Identificazione delle specifiche caratteristiche da estrarre / individuare automaticamente;
3. Arricchimento dei contenuti con metadati qualitativi e quantitativi connessi all'analisi linguistica (es. principali concetti, tipo di linguaggio usato, mood della conversazione, ...);
4. Sviluppo e test del sistema e sua integrazione con gli altri moduli MIRC.0.

Principali problematiche da affrontare:

1. Identificazione delle più promettenti combinazioni di metadati (pattern) di tipo semantico e quantitativo per l'identificazione di precise tipologie di utenti (cluster).

Tempi e modalità di realizzazione dell'intervento: Secondo la lista delle attività di seguito riportata.

Test da effettuare per misurare i parametri e tempi di realizzazione: L'efficacia dell'analisi linguistica sarà misurata principalmente dalla capacità / possibilità di automatizzare i processi (RPA). Un inteso utilizzo della RPA sarà infatti indice di un approccio un-supervised, automatico, e quindi in grado di "sgravare" la struttura da una serie di attività operative. Ovviamente perché ciò avvenga, sarà necessario aver individuato con precisione alcune peculiarità / caratteristiche dei vari Cluster.

Note descrittive generali: Se nell'OR4 il focus dell'analisi è rivolto alla totalità "indistinta" dei contenuti, la natura dell'analisi di questo OR dipende da alcune informazioni specifiche sulla persona coinvolta (o meglio sul tipo di persona) e sullo stato di avanzamento della pratica, tipicamente già disponibili all'interno dell'organizzazione (= informazioni a priori) che descrivono in modo più puntuale possibile i target dell'analisi. Per questo motivo, tale OR è anche strettamente connesso con lo studio della tipologia dell'interlocutore (profilo multidimensionale). L'assunto di base è che la conoscenza di alcune informazioni "a priori" possa indirizzare l'analisi ed ottenere informazioni più efficaci. A titolo esemplificativo, sapere a priori che il testo analizzato si riferisce ad una specifica fase del processo (ad esempio il primo contatto) ed il tipo debitore è "buono" o "cattivo", può rappresentare un valore aggiunto significativo nell'approccio all'analisi di text mining. In che misura? Scopo dell'OR è misurarlo effettuando la comparazione tra analisi "cieche" (a profilo libero) ed analisi con profilo pilotato (profilo multidimensionale) al fine di misurare l'efficacia.

L'approccio a questo OR vedrà il contributo di più metodologie:

- L'approfondimento / survey con gli stakeholder per raccogliere dal loro racconto alcune peculiarità / entità che possono essere utili all'analisi personalizzata
- L'adozione di tecniche / strumenti di ML per individuare in modo automatico eventuali pattern / concetti
- L'adozione della tecnologia semantica / NER, "potenziata" dallo sviluppo di specifiche regole di estrazione derivanti dai punti precedenti.

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE

OR5.1 Analisi del contesto

Obiettivo: Individuare una serie di metadati utili alla precisa clusterizzazione.

Organizzazione: La possibilità di creare un vero profilo "tailor made" rappresenta una sfida importante per il successo dell'intero progetto, in quanto più si ci si avvicinerà all'esatta definizione di un "soggetto tipo", tanto più le azioni di intervento potranno essere puntuali e mirate.

Tali informazioni potranno inoltre abilitare future forme evolute di dialogo automatico tra gli stakeholder, attraverso forme di chatbot o avatar integrati nelle future versioni della piattaforma. È evidente infatti, che maggiore è la capacità di distinguere specifiche fasi / aspettative, maggiore sarà la possibilità a tendere di gestire in automatico anche la complessa interazione tra operatore e debitore.

Nel corso dell'attività si definiranno, in un primo tempo, gli stakeholder e si organizzeranno, successivamente, riunioni e scambi di informazioni via web per definire le peculiarità / entità da estrarre dai testi. Il numero degli stakeholder potrà crescere anche in un secondo tempo, nel corso degli scambi e per conseguenza di questi. In aggiunta la co-proponente Fire, si considera che nel gruppo degli stakeholder possano trovarsi associazioni di categoria, rappresentanze professionali, associazioni di consumatori, etc.

Tempistica:

mese di inizio	1
mese di fine	18
durata (in mesi) dell'attività	18

Consegne:

1. Definizione e organizzazione dei gruppi di interesse (stakeholder): $T0 + 6$ mesi;
2. Output dell'analisi: $T0 + 18$ mesi.

OR5.2 Sviluppo del sistema

Obiettivo: Sviluppare il sistema di analisi specializzato secondo le indicazioni ottenute dall'attività OR5.1.

Organizzazione: Si adotteranno tecniche di ML per individuare i pattern / concetti in questione. L'attività si basa sugli strumenti già in possesso del proponente Expert System e su quelli sviluppati nel corso dell'OR4.

Tempistica:

mese di inizio	19
mese di fine	30
durata (in mesi) dell'attività	12

Consegne:

1. Versione 0: $T0 + 24$ mesi;
2. Versione finale: $T0 + 30$ mesi.

OR5.3 Test sul sistema sviluppato e feedback

Obiettivo: Verificare la capacità del sistema sviluppato di individuare cluster e informazioni significanti dai test

Organizzazione: I test saranno di due tipi:

- Automatici, di primo livello, realizzati attraverso il benchmark automatico dei risultati ottenuti su gruppi di controllo distinti individuati nel corpus di test Fire;
- Manuali, ovvero con intervento di operatori umani, attraverso l'analisi di test appositamente predisposti per questi test e il confronto con l'analisi svolta sugli stessi da una persona.

Tempistica: Per l'organizzazione dei test manuali occorre anticipare la parte di lavoro svolta dagli operatori umani; conseguentemente quest'attività inizia prima che lo sviluppo del sistema sia completo. Per i test propriamente detti, si considera una durata di 4 mesi, compresi i feedback sul sistema.

mese di inizio	24
mese di fine	34
durata (in mesi) dell'attività	11

Consegne:

1. Test predisposti: $T0 + 30$ mesi;
2. Ultimo feedback: $T0 + 34$ mesi.

MIRC.0

OR5.4 Integrazione con gli altri moduli

Obiettivo: Integrare quanto realizzato nel sistema MIRC.0.

Organizzazione: Tenendo conto del contesto dello sviluppo del sistema, l'integrazione del modulo al suo interno è da considerarsi standard.

Tempistica:

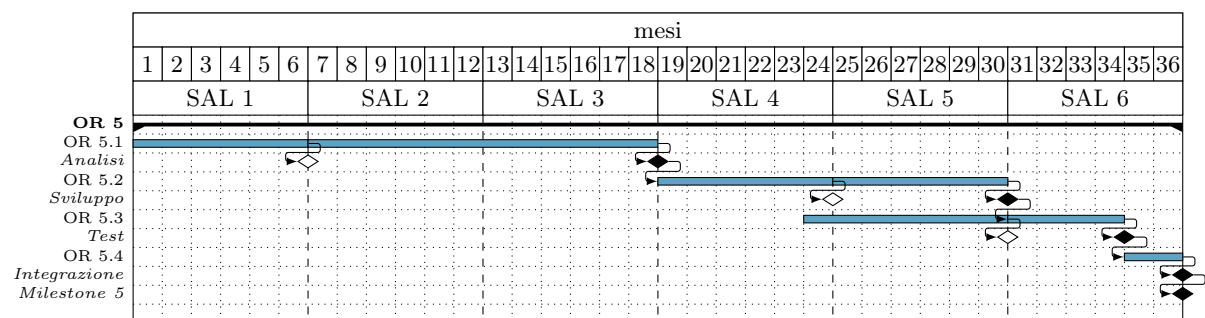
mese di inizio	35
mese di fine	36
durata (in mesi) dell'attività	2

Consegne:

1. Sistema integrato in MIRC.0: T0 + 36 mesi.

TEMPI DI REALIZZAZIONE

Durata (in mesi): 36



7.6. OR6 – Pilota del sistema NLP

DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO

Titolo dell'OR: Pilota del sistema NLP

Soggetto proponente: Expert System

Tipologia (RI/SS): SS

Tempistica dell'OR:	mese di inizio	1
	mese di fine	6
	durata (in mesi) dell'OR	6

Luoghi di svolgimento: Sede Expert System

Obiettivo: Sistema dimostrativo dei risultati dell'analisi dei contenuti non strutturati (testi). Premesso che l'integrazione di sistema NLP con altre componenti della piattaforma avverrà attraverso specifici Web services che garantiranno un'operatività tra moduli ICT (RPA), verrà sviluppata anche una semplice applicazione web in cui sarà possibile testare l'analisi di documenti di interesse da parte di personale non esperto in tecnologia di trattamento dei contenuti testuali. Lo studio dell'applicazione verrà svolto insieme agli altri Partner, ma in linea di massima il sistema permetterà:

- La selezione di un testo da analizzare
- L'elaborazione dello stesso
- La visualizzazione delle principali informazioni estratte / riconosciute
- La possibilità di esportare i risultati in forma digitale / strutturata

L'utilizzo del suddetto ambiente permetterà anche di intercettare, nel corso del tempo, eventuali scostamenti dei temi trattati (ad esempio in seguito al trattamento di nuovi domini informativi) e quindi permetterà di intercettare eventuali modifiche da apportare al "motore di analisi" integrato nella piattaforma.

Parametri che permettono di misurare il raggiungimento dell'obiettivo:

- Facilità di interazione da parte di personale non esperto;
- Estrazione di informazioni significanti generate dall'analisi Semantica.

Tipo di intervento:

- Interazione con operatori Fire (utenti del sistema) per raccogliere esigenze e feedback utilizzo;
- Interazione automatica / integrazione di sistema con altri moduli al fine di agevolare l'importazione / esportazione dei contenuti.

Principali problematiche da affrontare:

- Adottare una GUI (interfaccia grafica con l'utente) di semplice utilizzo e chiara;
- Automatizzare i processi al fine di rendere efficiente l'utilizzo del Pilot in modalità stand-alone.

Tempi e modalità di realizzazione dell'intervento: Secondo la lista delle attività di seguito riportata.

Test da effettuare per misurare i parametri e tempi di realizzazione: L'efficacia del PILOT sarà misurata attraverso delle interviste ai principali utenti della stessa, suddivisi per tipologia (operatore call center, analisti, manager, ...).

Note descrittive generali: Progettazione e sviluppo di un dimostratore del sistema di analisi NLP. Le potenzialità e caratteristiche dell'analisi linguistica profonda spesso non sono facilmente comprese da personale non-tecnico, ma magari esperto del dominio/business in cui si opera. È quindi molto utile fornire dei semplici strumenti per provare e comprendere le caratteristiche dell'analisi semantica, anche in modalità "stand alone" ed in ambiente dedicato, quindi direttamente accessibile, senza dovervi accedere attraverso altre piattaforme / strumenti. Quest'ambiente risulta utile anche per effettuare veloci verifiche / test / benchmarking e quindi ha anche valore per i tecnici più esperti.

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE

OR6.1 Progettazione interfaccia grafica

Obiettivo: Progettare una GUI adatta agli operatori di Fire

Organizzazione: In una prima fase si eseguirà il design dell'interfaccia basandosi su sistemi esistenti in ambienti simili e su interviste con il personale Fire. Si procederà successivamente alla programmazione dell'interfaccia e a test di funzionamento con utenti scelti a campione. Queste attività si svolgeranno parallelamente agli altri OR e si concluderanno prima che il sistema MIRC.0 sia pronto per ricevere l'interfaccia utente. A fine progetto, si proporrà una seconda campagna di prove con utenti scelti a campione, questa volta interfacciando l'interfaccia con il sistema di supporto alle decisioni completo.

Tempistica: L'attività sarà avviata a inizio progetto, fino a giungere alla realizzazione di un'interfaccia che simuli il sistema finale. Interrotta una volta giunti a questo punto, l'attività sarà riavviata al completamento del sistema di supporto alle decisioni, in modo da realizzare i test definitivi con utenti scelti a campione.

mese di inizio	1
mese di fine	30
durata (in mesi) dell'attività	30

Consegne:

1. Design della GUI: $T_0 + 6$ mesi;
2. Primi test con utenti: $T_0 + 18$ mesi;
3. Versione finale: $T_0 + 36$ mesi.

OR6.2 Sviluppo ambiente dimostrativo e di test continuo dell'analisi linguistica

Obiettivo: Far comprendere e utilizzare l'engine Cogito anche da personale non specializzato in NLP.

Organizzazione: Verrà sviluppata un'applicazione web per consentire l'analisi di documenti di interesse da parte di personale Fire. Il sistema dovrà essere in grado di intercettare eventuali scostamenti dei temi trattati (ad esempio in seguito al trattamento di nuovi domini informativi) per definire modifiche da apportare al "motore di analisi" integrato nella piattaforma.

Tempistica:

mese di inizio	13
mese di fine	30
durata (in mesi) dell'attività	18

Consegne:

1. Versione 0: $T_0 + 30$ mesi;
2. Versione finale: $T_0 + 36$ mesi.

OR6.3 Test con diverse categorie di utenti

Obiettivo: Verificare il livello di comprensione dei problemi che può essere ottenuto grazie all'impiego del sistema da parte di utenti non specializzati.

Organizzazione: Si svolgeranno due campagne di test col personale Fire, organizzato in categorie. La prima sarà organizzata a inizio progetto, per definire i parametri di riferimento dell'interfaccia, la seconda a fine progetto, per valutare il sistema.

Tempistica:

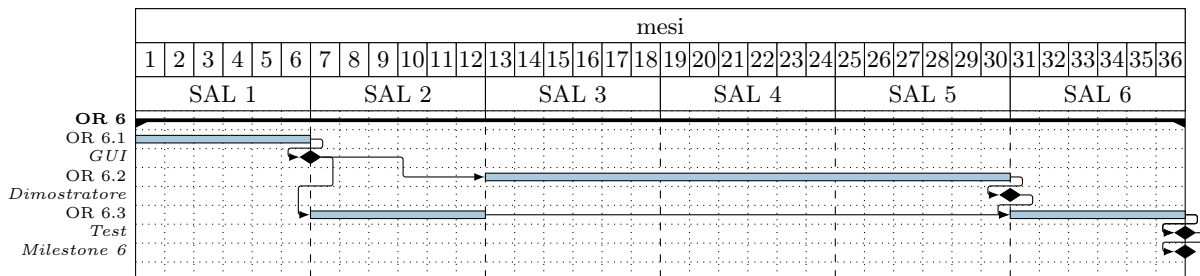
mese di inizio	7
mese di fine	36
durata (in mesi) dell'attività	30

Consegne:

1. Versione 0: T0 + 30 mesi;
2. Versione finale: T0 + 36 mesi.

TEMPI DI REALIZZAZIONE

Durata (in mesi): 36



7.7. OR7 – Data base

DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO

Titolo dell'OR: Data base

Soggetto proponente: Fire

Tipologia (RI/SS): RI

Tempistica dell'OR:	mese di inizio	1
	mese di fine	36
	durata (in mesi) dell'OR	36

Luoghi di svolgimento: Sede Fire Messina

Obiettivo: Organizzare l'insieme delle basi di dati utili alla realizzazione di MIRC.0. Le basi di dati cui il sistema dovrà attingere sono molteplici ed eterogenee. In una prima fase dell'OR occorrerà eseguirne una classificazione, valutare i formati esistenti, identificare eventuali evoluzioni utili per questi formati e definire i canali attraverso i quali il sistema avrà accesso ai dati. Una lista non esaustiva dei tipi di dati da consultare è la seguente:

- Norme e regolamenti;
- Contratti;

MIRC.0

- Statistiche;
- Proiezioni;
- Storico delle pratiche Fire.

In generale, ma soprattutto in relazione all'ultimo punto, occorrerà eseguire delle operazioni di pulitura preliminari rispetto all'impiego di algoritmi di ML.

Occorre inoltre che l'accesso ai dati sia tutelato da procedure di security verso l'esterno e da procedure di oscuramento interno, che dovranno avere luogo ogni qualvolta determinate informazioni debbano essere "dimenticate" dal sistema, per esempio a seguito del completamento di una trattativa.

Fire è attualmente certificata ISO 27001 per la sicurezza dei dati e esegue al proprio interno procedure di oscuramento previste dai rapporti contrattuali. Nel corso di questo OR si dovrà comunque verificare che MIRC.0 non introduca criticità relativamente a questi standard, implementando eventualmente i necessari algoritmi di gestione e controllo.

Parametri che permettono di misurare il raggiungimento dell'obiettivo:

1. Numero di basi dati accessibili;
2. Qualità dei dati rispetto ai sistemi da utilizzare;
3. Qualità delle procedure di sicurezza;
4. Qualità delle procedure di oscuramento.

Tipo di intervento:

1. Istituzione di un gruppo di lavoro composto da personale Fire che piloti le azioni sui dati; questo gruppo sarà denominato "TeamOR7", per distinguerlo dai corrispettivi negli altri OR;
2. Organizzazione di riunioni di lavoro tra operatori Fire per raccogliere esigenze sui dati;
3. Definizione delle base dati cui accedere e dei relativi canali;
4. Acquisti, noleggi, licenze relativi a tali basi;
5. Organizzazione e realizzazione del sistema per la pulizia dei dati;
6. Organizzazione e implementazione delle procedure di sicurezza;
7. Organizzazione e implementazione delle procedure di oscuramento.

Principali problematiche da affrontare: Gli interventi sopra enumerati contengono una comune difficoltà, dovuta all'eterogeneità dei dati da trattare, che potrà essere mitigata dall'esperienza del proponente nel settore di riferimento.

Il problema principale sui dati è l'individuazione, catalogazione e classificazione di tutte le informazioni provenienti dalle varie fonti, che possono avere rilevanza nella generazione dei processi decisionali. Molte di queste informazioni risiedono all'interno di database pubblici, altre si possono acquisire attraverso l'anonimizzazione e l'aggregazione statistica dei dati prodotti durante l'attività di gestione dei crediti; altre risiedono negli scambi verbali che si hanno durante il dialogo collector-debitore e sono trattati dagli OR4, 5 e 6 del "sistema dialogo".

Tempi e modalità di realizzazione dell'intervento: Secondo la lista delle attività di seguito riportata

Test da effettuare per misurare i parametri e tempi di realizzazione: Per ciascuna delle basi dati che si sanno identificate durante la prima fase dell'OR, si potranno realizzare, a fine progetto, dei test relativi ai tre ultimi punti della lista precedente.

Note descrittive generali: Il progetto intende trattare il problema della tutela del credito in modo innovativo, tenendo conto delle esigenze dell'interlocutore e prospettando il ricorso a percorsi modulari sulla base di queste esigenze. Affinché quest'ambizione possa essere soddisfatta, è necessario poter accedere e integrare nel sistema un gran numero di informazioni, alcune – specifiche al contesto – già associate al processo decisionale attualmente messo in atto, altre – da definire – tali da consentire la corretta rappresentazione delle esigenze e delle opportunità dell'interlocutore.

Il problema di quali siano queste informazioni è totalmente aperto in questa fase, e dovrà essere definito nel corso dell'OR. Si può immaginare tuttavia che l'accesso a informazioni statistiche e proiezioni nazionali possa essere di aiuto nel valutare l'effettiva possibilità di un pagamento o l'effettivo valore di una dilazione.

Le scelte che portano ad identificare soluzioni idonee a risolvere problematiche debitorie sono molteplici. I fattori principali su cui si basano sono: la tipologia del credito (prestito personale, mutuo, utenza, scopertura di conto,...); l'anzianità del debito (Aging); la condizione economica dell'interlocutore; la volontà dell'interlocutore; le regole di definizione indicate dal titolare del credito, vincolate da proprie policy e al contratto che sta alla base del credito stesso. Si tratta di un sistema estremamente variegato ed entropico. Poter garantire equità di comportamento e trattamento nei confronti di tutti i soggetti che si trovano in difficoltà, poter applicare in maniera quanto più obiettiva e customizzata tutte le regole che stanno alla base del processo che porta alla definizione di una proposta transattiva, è un processo complesso che si basa molto sulla formazione delle risorse addette al rapporto con i soggetti debitori, e prevede processi operativi articolati e differenziati per tipologia di credito.

Per quanto si possano applicare principi di corretta process governance e interventi formativi, il momento nel quale si confrontano collector e debitore risulta fortemente caratterizzato dall'iterazione a due vie tra questi due soggetti. Quest'elemento fortemente personale incide in maniera significativa nell'individuazione della soluzione transattiva, risultando in alcuni casi anche limitante.

Riuscire ad ingegnerizzare sia le conoscenze delle regole di gestione tipiche di ciascun prodotto, sia le conoscenze derivanti dall'analisi della dinamica con la quale si svolge l'interlocuzione collector-debitore, è la base per il raggiungimento dell'obiettivo perseguito dai tre OR del "sistema scelte". Per gli scopi del presente progetto, le azioni necessarie a tal fine sono state appunto raggruppate negli OR7, 8 e 9, intendendo in tal modo separare:

- (OR7) l'analisi della natura e della struttura dei dati a priori sui quali si basano le scelte in esame;
- (OR8) l'analisi della componente umana, secondo l'organizzazione interna di Fire, le modalità con le quali queste risorse interagiscono tra di loro e con gli interlocutori esterni, le modalità con le quali le stesse risorse dovrebbero attingere informazioni e ottenere supporto da MIRC.0, e le modalità con le quali MIRC.0 dovrà attingere informazioni da loro;
- (OR9) i test di funzionamento di MIRC.0 e del sistema delle risorse umane quando supportate da MIRC.0.

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE

OR7.1 Definizione delle basi dati da utilizzare

Obiettivo: Definire l'insieme delle basi dati da utilizzare e organizzare i canali di accesso

Organizzazione: In primo luogo verrà istituito il gruppo di lavoro "TeamOR7" composto da personale Fire cui siano note le problematiche sui dati. Questo gruppo si occuperà di raccogliere l'informazione all'interno della ditta sulle tipologie di dati di cui è utile l'acquisizione, aggiungendo a quelle comunemente usate secondo le procedure attuali altre che possono essere suggerite dall'esperienza del personale.

L'attività serve a specificare quali siano le basi dati di interesse per il sistema MIRC.0, tenendo conto del fatto che esso deve fornire risposte utili valutate secondo molteplici misure, valide in diversi contesti. Ad esempio possiamo citare le leggi in vigore, i regolamenti, i contratti, etc., ma anche altre informazioni, eventualmente statistiche, sociologiche o soggettive, che permettono di svolgere pratiche di recupero che tengano conto delle reali esigenze dell'interlocutore.

MIRC.0

Tempistica:

mese di inizio	1
mese di fine	12
durata (in mesi) dell'attività	12

Consegne:

1. Elenco basi oggettive e accesso a queste basi: $T0 + 6$ mesi;
2. Elenco basi statistiche o altro e accesso a queste basi: $T0 + 12$ mesi.

OR7.2 Pulizia dei dati

Obiettivo: Definire appositi filtri affinché le basi cui si deve accedere siano prive di ridondanze ed errori che possono inficiare il funzionamento del sistema

Organizzazione: Il problema si presenta soprattutto per il data base Fire, ma anche per altre basi provenienti da sorgenti diverse, come documenti generati da associazioni di categoria, uffici studi, analisti, etc.

Tempistica:

mese di inizio	7
mese di fine	18
durata (in mesi) dell'attività	12

Consegne:

1. Organizzazione delle procedure per la base dati Fire $T0 + 12$ mesi;
2. Sistema di pulizia per basi generiche: $T0 + 18$ mesi.

OR7.3 Sicurezza dei dati

Obiettivo: Definire e adottare i sistemi di sicurezza dei dati

Organizzazione: Si farà riferimento a sistemi esistenti, avendo cura di scegliere quelli più adatti al contesto.

Tempistica:

mese di inizio	13
mese di fine	30
durata (in mesi) dell'attività	18

Consegne:

1. Scelta del sistema $T0 + 21$ mesi;
2. Sistema funzionante: $T0 + 30$ mesi.

OR7.4 Oscuramento dei dati

Obiettivo: Definire e adottare i sistemi di oscuramento dei dati

Organizzazione: Si farà riferimento a sistemi esistenti, avendo cura di scegliere quelli più adatti al contesto.

Tempistica:

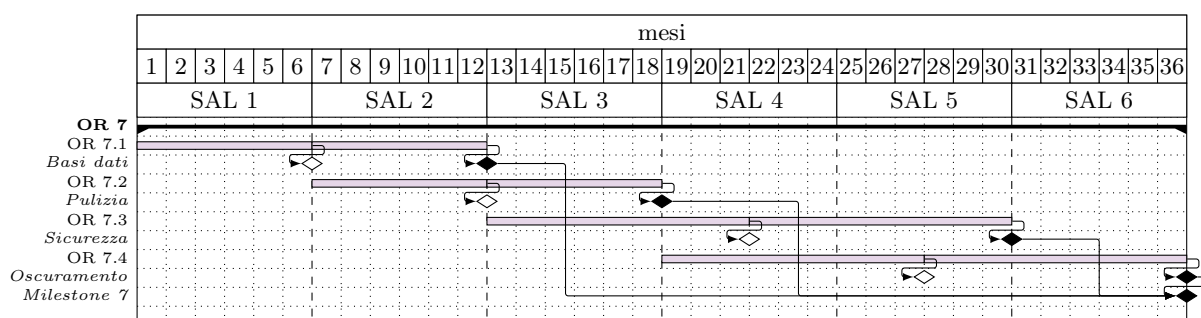
mese di inizio	19
mese di fine	36
durata (in mesi) dell'attività	18

Consegne:

1. Scelta del sistema T0 + 27 mesi;
2. Sistema funzionante: T0 + 36 mesi.

TEMPI DI REALIZZAZIONE

Durata (in mesi): 36

**7.8. OR8 – Sistema utente****DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO**

Titolo dell'OR: Sistema utente

Soggetto proponente: Fire

Tipologia (RI/SS): SS

Tempistica dell'OR:	mese di inizio	1
	mese di fine	31
	durata (in mesi) dell'OR	31

Luoghi di svolgimento: Sede Fire Messina

Obiettivo: Definire il sistema utente esistente e organizzare il sistema utente del progetto.

Ci si riferisce a un "sistema" utente piuttosto che a un insieme di utenti in quanto MIRC.0 dovrà fornire supporto a decisioni che vengono prese a più livelli su una stessa pratica. Pertanto è necessario, da una parte, trasferire l'esperienza di operatori di diverso livello verso MIRC.0, dall'altra fare in modo che MIRC.0 possa produrre supporti con informazioni aggregate in modo variabile secondo il livello decisionale al quale si rivolge.

La definizione e l'organizzazione del sistema utente che può essere ottenuta grazie all'innovazione di prodotto introdotta da MIRC.0, configurano, infine, una fondamentale innovazione di processo all'interno dell'azienda, basata sulla possibilità fornita da MIRC.0 di sviluppi basati sui dati.

La strutturazione attuale delle risorse umane operative di Fire si configura su due livelli:

1. "Collector", ovvero operatori che sono a contatto con il debitore, suddivisi in categorie quali:

MIRC.0

- “Phone collector” (contatto telefonico);
 - “Home collector” (contatto personale diretto);
 - “Legal collector” (contatto legale);
2. “Production manager”, che sono in contatto con i collector, ma non con il debitore.

Obiettivo dell’OR è l’ingegnerizzazione delle conoscenze di questi gruppi di risorse e la definizione del nuovo modo di funzionare che potrà essere messo in atto grazie a MIRC.0. Per un corretto sviluppo di questa transizione è necessario un intenso scambio di informazioni e di opinioni tra il personale addetto al progetto MIRC.0 e il personale operativo. L’OR si articola dunque in attività nel corso delle quali queste interazioni sono organizzate e messe in atto, per giungere infine a una visione complessiva il più che possibile esatta dei metodi da implementare tramite MIRC.0.

Parametri che permettono di misurare il raggiungimento dell’obiettivo:

1. Numero delle tipologie di pratiche per le quali è stato definito un processo che può essere assistito dal sistema di supporto delle decisioni MIRC.0;
2. Numero (percentuale ponderata secondo criteri da definire) di interventi umani all’interno di queste pratiche per i quali il sistema MIRC.0 può essere di supporto;

Tipo di intervento:

1. Istituzione di un gruppo di lavoro composto da personale Fire che piloti le azioni di ingegnerizzazione delle conoscenze; questo gruppo sarà denominato “TeamOR8”, per distinguerlo dai corrispettivi negli altri OR;
2. Organizzazione di riunioni di lavoro tra operatori Fire per identificare le procedure effettive attuali, tenendo conto – come possibile – del contesto delle interazioni umane con gli interlocutori;
3. Definizione dei requisiti del sistema MIRC.0, organizzati secondo livelli di intervento di importanza crescente (vedi sotto).

Principali problematiche da affrontare: Problematiche tipiche della knowledge engineering (KE) applicata a società di servizio.

Tempi e modalità di realizzazione dell’intervento: Secondo la lista delle attività di seguito riportata

Test da effettuare per misurare i parametri e tempi di realizzazione: Quanto fatto in questo OR potrà essere convalidato in sede di validazione del sistema MIRC.0, ovvero nel corso dell’OR9.

Note descrittive generali: L’organizzazione della conoscenza è correlata alla complessità del sistema nel quale essa si è sviluppata. Occorre quindi identificare i meccanismi specifici del contesto Fire prima di poter pervenire a rendere utilizzabile in maniera ingegneristica le fonti di conoscenza che risiedono nei data base trattati dall’OR7 e nel comportamento degli operatori, oggetto di questo OR.

Ogni esperto umano, in quanto tale, si confronta abitualmente con informazioni ambigue, valutazioni approssimative, variabili non completamente definite o identificabili. Il rapporto tra la qualità della proiezione o della predizione che consente di operare la scelta in modo razionale e i costi per la massimizzazione di questa qualità è spesso a vantaggio di una soluzione basata su informazioni non ottimali. La scelta operata da un esperto è dunque un’atto di compromesso, ritenuto soddisfacente secondo – appunto – l’esperienza del soggetto.

Nel contesto Fire, se è possibile che un sistema di supporto delle decisioni migliori certe possibili scelte, è anche vero che l’ingegnerizzazione della componente umana del profilo dell’interlocutore resta difficilmente oggettivabile. Certamente, un sistema ICT ha la possibilità di:

- accedere rapidamente ai dati e non commettere errori relativamente all'informazione a priori sulla pratica;
- eliminare rapidamente opzioni molto probabilmente infruttuose sulla base dell'esperienza generata dal ML;
- analizzare il problema in modo differenziato, con diversi livelli di astrazione.

Quello che è necessario aggiungere al sistema rispetto a queste caratteristiche tipicamente collegate alla sua natura informatica, deve essere creato attraverso le interviste col personale. Ci riferiamo in particolare alla possibilità di simulare expertise diverse, provenienti da operatori diversi, ma classificate in MIRC.0 grazie alle attività di questo OR. Le interviste saranno quindi prevalentemente mirate a:

- raccogliere regole euristiche;
- creare uno schema di rappresentazione della conoscenza proprio dei diversi gruppi di risorse;
- identificare possibili segmentazioni dei processi.

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE

OR8.1 Metodologie di base

Obiettivo: Istituzione del gruppo di lavoro "TeamOR8" e definizione del calendario delle interviste, secondo i diversi gruppi di personale e tematiche da affrontare nelle consultazioni

Organizzazione: Scopo dell'attività è il setup dell'OR, che comporta l'organizzazione di scambi diretti e indiretti con il personale. Per quanto riguarda i primi, si tratta di prevedere riunioni in piccoli gruppi di lavoro per la trasmissione delle conoscenze verso il sistema MIRC.0. Relativamente agli scambi indiretti, si tratta di organizzare sistemi basati su tecnologie web, tipo blog o simili, per raccogliere informazioni a distanza in modo ordinato, secondo schemi che dovranno essere definiti nel corso di quest'attività.

Tempistica:

mese di inizio	1
mese di fine	4
durata (in mesi) dell'attività	4

Consegne:

1. Organizzazione dell'OR: $T_0 + 4$ mesi;

OR8.2 Analisi della conoscenza dei phone collector

Obiettivo: Svolgimento delle azioni programmate nel corso di OR8.1 in relazione alla conoscenza dei phone collector.

Organizzazione: Secondo quanto previsto in OR8.1

Tempistica: Si prevedono output semestrali dei risultati di queste analisi

mese di inizio	5
mese di fine	28
durata (in mesi) dell'attività	24

MIRC.0

Consegne:

1. Output 1: $T_0 + 10$ mesi;
2. Output 2: $T_0 + 16$ mesi;
3. Output 3: $T_0 + 22$ mesi;
4. Output finale: $T_0 + 28$ mesi.

OR8.3 Analisi della conoscenza degli home collector

Obiettivo: Svolgimento delle azioni programmate nel corso di OR8.1 in relazione alla conoscenza degli home collector.

Organizzazione: Secondo quanto previsto in OR8.1

Tempistica: Si prevedono output semestrali dei risultati di queste analisi

mese di inizio	6
mese di fine	29
durata (in mesi) dell'attività	24

Consegne:

1. Output 1: $T_0 + 11$ mesi;
2. Output 2: $T_0 + 17$ mesi;
3. Output 3: $T_0 + 23$ mesi;
4. Output finale: $T_0 + 29$ mesi.

OR8.4 Analisi della conoscenza dei legal collector

Obiettivo: Svolgimento delle azioni programmate nel corso di OR8.1 in relazione alla conoscenza dei legal collector.

Organizzazione: Secondo quanto previsto in OR8.1

Tempistica: Si prevedono output semestrali dei risultati di queste analisi

mese di inizio	7
mese di fine	30
durata (in mesi) dell'attività	24

Consegne:

1. Output 1: $T_0 + 12$ mesi;
2. Output 2: $T_0 + 18$ mesi;
3. Output 3: $T_0 + 24$ mesi;
4. Output finale: $T_0 + 30$ mesi.

OR8.5 Analisi della conoscenza dei production manager

Obiettivo: Svolgimento delle azioni programmate nel corso di OR8.1 in relazione alla conoscenza dei production manager.

Organizzazione: Secondo quanto previsto in OR8.1

Tempistica: Si prevedono output semestrali dei risultati di queste analisi

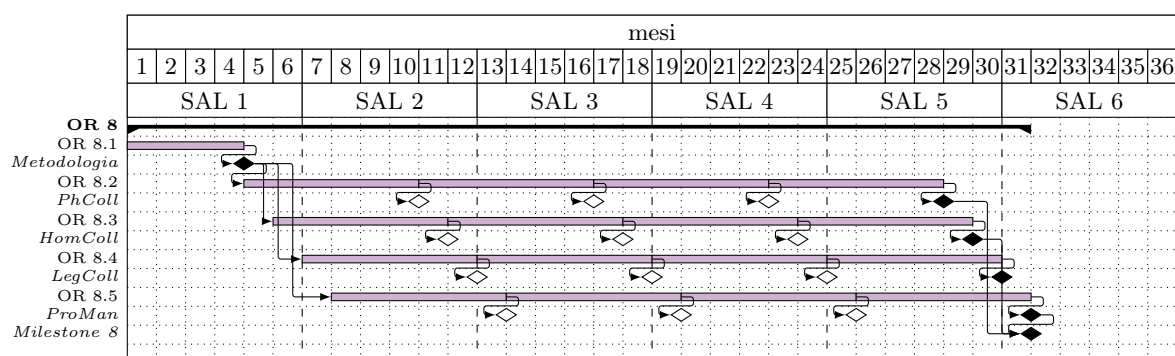
mese di inizio	8
mese di fine	31
durata (in mesi) dell'attività	24

Consegne:

1. Output 1: T0 + 13 mesi;
2. Output 2: T0 + 19 mesi;
3. Output 3: T0 + 25 mesi;
4. Output finale: T0 + 31 mesi.

TEMPI DI REALIZZAZIONE

Durata (in mesi): 31



7.9. OR9 – Valutazione del sistema MIRC.0

DESCRIZIONE DELL'OBIETTIVO REALIZZATIVO

Titolo dell'OR: Valutazione del sistema MIRC.0

Soggetto proponente: Fire

Tipologia (RI/SS): SS

Tempistica dell'OR:	mese di inizio	10
	mese di fine	36
	durata (in mesi) dell'OR	27

Luoghi di svolgimento: Sede Fire Messina

Obiettivo: Valutare l'impiego del sistema MIRC.0 e i suoi effetti sull'organizzazione del lavoro.

Parametri che permettono di misurare il raggiungimento dell'obiettivo:

1. Coerenza dei suggerimenti forniti dal sistema con le azioni che sarebbero state intraprese da operatori non supportati dal sistema (percentuale di successi);
2. Valutazione del supporto ottenuto da parte degli operatori dei diversi livelli;
3. Possibilità di semplificare l'approccio alle pratiche e ridistribuire le risorse umane in funzione della complessità delle stesse (valorizzazione dell'apporto delle risorse ai risultati);
4. Feedback da sistemi di valutazione esterni (come associazioni di consumatori).

Tipo di intervento:

1. Istituzione di un gruppo di lavoro composto da personale Fire che piloti le azioni di valutazione del sistema MIRC.0; questo gruppo sarà denominato "TeamOR9", per distinguerlo dai corrispettivi negli altri OR;
2. Organizzazione dei sistemi di controllo, distinti in semi-automatici, con simulazione e diretti:
 - Si intende semi-automatico un controllo eseguito applicando MIRC.0 a un insieme di pratiche di controllo scelte appositamente per valutare la capacità del sistema in condizioni critiche. Le pratiche saranno estratte dalla base dati storica (e pertanto non utilizzate in fase di apprendimento), secondo criteri che verranno stabiliti nella fase iniziale dell'OR.
 - Si intende per controllo con simulazione una procedura nella quale due operatori Fire dialogano. L'operatore A è assistito da MIRC.0 e non sa di giocare in una simulazione, mentre l'operatore B gioca un ruolo prestabilito per la simulazione.
 - Si intende infine per analisi diretta della risposta, l'analisi della risposta proposta da MIRC.0 in condizioni di funzionamento reale, eseguita in tempo reale o a posteriori da personale specializzato.
3. Feedback dei risultati ottenuti dai diversi tipi di controlli;
4. Interazione con agenti esterni (per esempio associazioni di consumatori) per valutare l'effetto del sistema sugli interlocutori, se non su basi reali e statistiche, almeno sulla base delle proposte effettuate da MIRC.0 in seno alle predette simulazioni.

Principali problematiche da affrontare: Organizzazione di test che coinvolgono tecnologie ICT e risorse umane.

Tempi e modalità di realizzazione dell'intervento: Secondo la lista delle attività di seguito riportata

Test da effettuare per misurare i parametri e tempi di realizzazione: L'OR ha lo scopo di convalidare la realizzazione finale del progetto. I test che permettono di valutare l'efficacia delle azioni svolte riguardano il confronto tra il prodotto ottenuto a monte dei test e il prodotto a valle.

Note descrittive generali: L'OR comporta quattro famiglie di prove, tre da svolgere all'interno di Fire e una in relazione a interlocutori esterni. L'organizzazione e lo svolgimento esatto delle prove verranno definiti in fase di progetto, secondo i risultati ottenuti negli OR precedenti.

ELENCO DELLE ATTIVITÀ DELL'OBIETTIVO REALIZZATIVO E RELATIVA DESCRIZIONE

OR9.1 Metodologie di base

Obiettivo: Istituzione del gruppo di lavoro "TeamOR9"; definizione dei test (o delle metodologie per definire i test) e definizione del calendario dei lavori.

Organizzazione: Scopo dell'attività è il setup dell'OR, che comporta l'organizzazione dei test sopra descritti.

Tempistica:

mese di inizio	10
mese di fine	11
durata (in mesi) dell'attività	2

Consegne:

1. Organizzazione dell'OR: $T0 + 11$ mesi;

OR9.2 Test semi-automatici

Obiettivo: Scelta delle registrazioni da impiegare come gruppo di controllo. Organizzazione e svolgimento dei test.

Organizzazione: I test verranno definiti sulla base di quanto ottenuto dalle analisi dell'OR8. Conseguentemente sarà possibile specificare le caratteristiche dei dialoghi che devono essere selezionati come gruppo di controllo. Una volta definite queste caratteristiche, si potranno fornire indicazioni ai partner per eseguire una selezione automatica dei testi di interesse.

Tempistica:

mese di inizio	12
mese di fine	34
durata (in mesi) dell'attività	23

Consegne:

1. Definizione dei test: $T0 + 24$ mesi;
2. Test pronti per l'esecuzione: $T0 + 28$ mesi;
3. Fine test: $T0 + 34$ mesi;

OR9.3 Test con simulazione

Obiettivo: Definizione degli scenari da simulare. Organizzazione e svolgimento dei test

Organizzazione: Gli scenari da simulare verranno definiti sulla base di quanto ottenuto dalle analisi dell'OR8. Del personale sarà istruito per giocare il ruolo dell'interlocutore nelle simulazioni.

Tempistica:

mese di inizio	18
mese di fine	34
durata (in mesi) dell'attività	17

Consegne:

1. Definizione dei test: $T0 + 24$ mesi;
2. Test pronti per l'esecuzione: $T0 + 28$ mesi;
3. Fine test: $T0 + 34$ mesi;

MIRC.0

OR9.4 Analisi diretta

Obiettivo: Istruzione agli operatori, organizzazione e svolgimento dei test.

Organizzazione: L’istruzione da fornire agli operatori riguarda le metodologie e i criteri con i quali si assegneranno le valutazioni delle risposte di MIRC.0 nelle diverse circostanze.

Tempistica:

mese di inizio	18
mese di fine	34
<hr/>	
durata (in mesi) dell’attività	17

Consegne:

- 1. Istruzione operatori: T0 + 24 mesi;
- 2. Test pronti per l’esecuzione: T0 + 28 mesi;
- 3. Fine test: T0 + 34 mesi;

OR9.5 Interazione con agenti esterni

Obiettivo: Ottenere il parere di agenti esterni su quanto può essere ottenuto con MIRC.0.

Organizzazione: In un primo tempo occorrerà identificare gli interlocutori esterni. Successivamente si potranno stabilire le modalità di dialogo e di notazione del sistema. L’organizzazione dell’attività è da specificare in funzione degli accordi ottenuti in tal senso.

Tempistica:

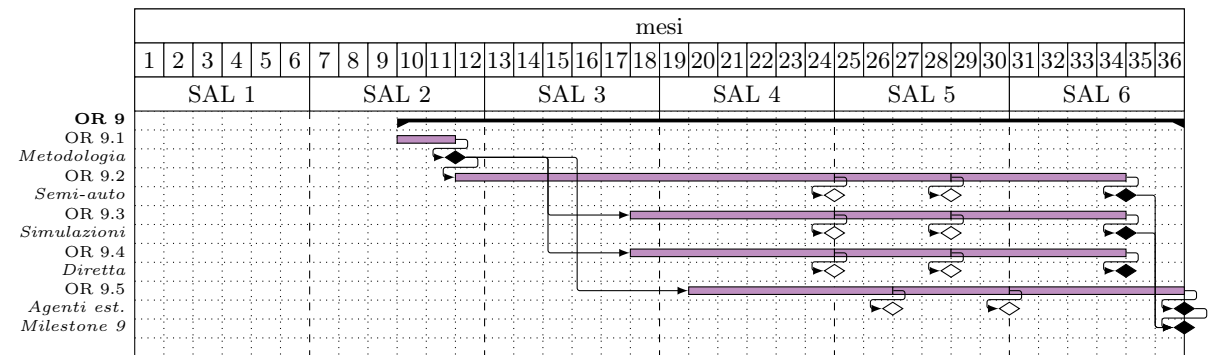
mese di inizio	20
mese di fine	36
<hr/>	
durata (in mesi) dell’attività	17

Consegne:

- 1. Definizione dei contatti: T0 + 26 mesi;
- 2. Test pronti per l’esecuzione: T0 + 30 mesi;
- 3. Fine test: T0 + 36 mesi;

TEMPI DI REALIZZAZIONE

Durata (in mesi): 27



8 RISULTATO INTERMEDIO ATTESO DEL PROGETTO

I risultati attesi a 18 mesi dall'inizio del progetto sono elencati qui di seguito.

- OR1.1 Identificazione dei dati, etichettatura dei dati e tipologie di analisi
 1. Prima identificazione dei dati (strutturati) disponibili e loro etichettatura: $T0 + 6$ mesi
 2. Seconda identificazione dei dati (strutturati) disponibili e identificazione di più etichettature utili al ML: $T0 + 12$ mesi
- OR1.2 Algoritmi di ML non supervisionato
 1. Prima identificazione degli algoritmi/strumenti più utili: $T0 + 9$ mesi
 2. Prima applicazione degli algoritmi/strumenti ai dati: $T0 + 18$ mesi
- OR1.3 Algoritmi di ML supervisionato
 1. Prima identificazione degli algoritmi/strumenti più utili: $T0 + 9$ mesi
 2. Prima applicazione degli algoritmi/strumenti ai dati etichettati: $T0 + 18$ mesi
- OR2.1 Creazione data base integrato
 1. Versione 1 del data base integrato: $T0 + 6$ mesi;
 2. Versione 2: $T0 + 18$ mesi;
- OR2.2 Principali determinanti del default
 1. Prima serie di determinanti del default: $T0 + 12$ mesi
- OR2.3 Principali Determinanti della LGD
 1. Prima serie di determinanti della LGD: $T0 + 18$ mesi;
- OR3.1 Definizione degli esempi per convalida e verifica sperimentale dei dimostratori
 1. Identificazione “esempi” per convalida e verifica sperimentale dei dimostratori: $T0 + 12$ mesi
- OR3.2 Sviluppo e applicazione dei dimostratori
 1. Definizione GUI dei dimostratori: $T0 + 12$ mesi ;
- OR4.1 Creazione Corpus file audio e documenti taggati
 1. Versione 1 del corpus files Audio e del corpus documenti: $T0 + 6$ mesi;
 2. Versione 2: $T0 + 12$ mesi;
- OR4.2 Riconoscimento file audio (ASR)
 1. Primo Corpus file trascritti (circa 10%): $T0 + 6$ mesi

MIRC.0

2. Secondo Corpus file trascritti (circa 50%): T0 + 12 mesi
- OR4.3 Analisi linguistica profonda dei testi
 1. Versione 0: T0 + 12 mesi;
 2. Versione 1: T0 + 24 mesi;
- OR5.1 Analisi del contesto
 1. Definizione e organizzazione dei gruppi di interesse (stakeholder): T0 + 6 mesi;
 2. Output dell'analisi: T0 + 18 mesi.
- OR5.2 Sviluppo del sistema
Nessun risultato atteso nei primi 18 mesi.
- OR5.3 Test sul sistema sviluppato e feedback
Nessun risultato atteso nei primi 18 mesi.
- OR5.4 Integrazione con gli altri moduli
Nessun risultato atteso nei primi 18 mesi.
- OR6.1 Progettazione interfaccia grafica
 1. Design della GUI: T0 + 6 mesi;
- OR6.2 Sviluppo ambiente dimostrativo e di test continuo dell'analisi linguistica
Nessun risultato atteso nei primi 18 mesi.
- OR6.3 Test con diverse categorie di utenti
Nessun risultato atteso nei primi 18 mesi.
- OR7.1 Definizione delle basi dati da utilizzare
 1. Elenco basi oggettive e accesso a queste basi: T0 + 6 mesi;
 2. Elenco basi statistiche o altro e accesso a queste basi: T0 + 12 mesi.
- OR7.2 Pulizia dei dati
 1. Organizzazione delle procedure per la base dati Fire T0 + 12 mesi;
 2. Sistema di pulizia per basi generiche: T0 + 18 mesi.
- OR7.3 Sicurezza dei dati
Nessun risultato atteso nei primi 18 mesi.
- OR7.4 Oscuramento dei dati
Nessun risultato atteso nei primi 18 mesi.
- OR8.1 Metodologie di base
 1. Organizzazione dell'OR: T0 + 4 mesi;
- OR8.2 Analisi della conoscenza dei phone collector
 1. Output 1: T0 + 10 mesi;
 2. Output 2: T0 + 16 mesi;

- OR8.3 Analisi della conoscenza degli home collector
 1. Output 1: $T_0 + 11$ mesi;
 2. Output 2: $T_0 + 17$ mesi;
- OR8.4 Analisi della conoscenza dei legal collector
 1. Output 1: $T_0 + 12$ mesi;
 2. Output 2: $T_0 + 18$ mesi;
- OR8.5 Analisi della conoscenza dei production manager
 1. Output 1: $T_0 + 13$ mesi;
- OR9.1 Metodologie di base
 1. Organizzazione dell'OR: $T_0 + 11$ mesi;
- OR9.2 Test semi-automatici

Nessun risultato atteso nei primi 18 mesi.
- OR9.3 Test con simulazione

Nessun risultato atteso nei primi 18 mesi.
- OR9.4 Analisi diretta

Nessun risultato atteso nei primi 18 mesi.
- OR9.5 Interazione con agenti esterni

Nessun risultato atteso nei primi 18 mesi.

9 RISULTATO FINALE ATTESO DEL PROGETTO

Il risultato finale è un sistema di supporto alle decisioni, MIRC.0, di cui si potrà valutare la rispondenza con i parametri di progetto. Le caratteristiche e le prestazioni che si vogliono ottenere sono state descritte nel §5.1 della II^a Parte del presente documento. Riprendendole qui di seguito, notiamo che MIRC.0 dorà:

Caratteristiche

1. Acquisire esperienza dai dati storici sulle pratiche di recupero crediti;
2. Acquisire esperienza durante lo svolgimento dei propri compiti sulle pratiche attuali;
3. Identificare l'interlocutore secondo un'apposita classificazione, che tenga conto – tra l'altro – delle condizioni attuali e future di stress finanziario che possono averlo colpito;
4. Comprendere il problema specifico posto all'interlocutore dalla pratica in corso;
5. Individuare il migliore percorso, integrando informazioni provenienti da tutte le fonti necessarie allo scopo, per ottenere la soluzione della controversia, secondo il profilo dell'interlocutore e a partire da una qualunque condizione a priori – dove per “migliore” si intende un percorso che tenga conto in modo responsabile del peso sociale dell'operazione;
6. Scambiare informazioni con l'interlocutore in modo da procedere secondo percorsi ottimali, via via aggiornati sulla base di queste informazioni, fino a risolvere la pratica.

Si osservi che il secondo punto della lista sarà perseguito nel corso del progetto, ma non si pretende di arrivare a un sistema industriale perfettamente autonomo in tal senso. Nella versione che si intende realizzare, MIRC.0 dovrà essere aggiornato (non automaticamente) con opportuna frequenza (per esempio trimestrale, o semestrale) in modo da incorporare l'esperienza contenuta nelle pratiche svolte nel periodo di tempo intercorso dall'ultimo aggiornamento.

Anche relativamente all'ultimo punto segnaliamo che si tratta di un obiettivo da perseguire in proiezione su tempi più lunghi di quelli del progetto, lo scambio con l'interlocutore potendo avvenire in un primo tempo esclusivamente attraverso un'interfaccia dedicata, tramite la quale l'operatore (“collector”) acquisisce informazioni e suggerimenti da MIRC.0, mentre dialoga in modo abituale con l'interlocutore. L'aggiornamento del percorso delle scelte sarà quindi, in questa configurazione, realizzato solo attraverso le domande poste al sistema dall'operatore.

Come indicato nello stesso paragrafo sopra citato, ci si aspetta di ottenere che MIRC.0 sia capace di prestazioni di tre livelli diversi: gestione del rischio, sistema di supporto alle decisioni e interfaccia conversazionale. Di queste, si ritiene che le prime due dovranno essere raggiunte, mentre la terza costituisce un obiettivo a più lungo termine.

Riportiamo quindi qui le prime due, come appaiono nel paragrafo citato, limitandoci agli aspetti che dovranno servire per validare l'attività svolta.

Prestazione 1: gestione del rischio MIRC.0 dovrà migliorare notevolmente la capacità predittiva del rischio e di rendimento atteso da ogni pratica e, al contempo, migliorare l'identificazione della modalità di recupero ottimale anche per l'interlocutore. Questo miglioramento può consentire di individuare le pratiche a più alta probabilità di soddisfazione del credito e concentrare l'utilizzo di risorse umane su queste ultime, permettendo all'impresa di fare leva sulla esistente forza lavoro per incrementare l'attività di gestione. Una migliore capacità predittiva del rischio può anche consentire di spingere al di là dei

livelli oggi accettabili la trattativa con interlocutori ritenuti affidabili (per esempio, individui sottoposti a momentanea pressione economica comprovata dall'incrocio di dati).

Prestazione 2: sistema di supporto alle decisioni In quanto sistema esperto interattivo, MIRC.0 può risolvere o evitare rischi di inefficienza, perdita di informazione, riduzione delle opzioni che possono essere generati nel corso delle interazioni tra operatori umani, oggi necessarie allo svolgimento delle pratiche.

In un primo tempo, il sistema si limiterà a interagire solo con gli operatori umani, mediante un'interfaccia appositamente progettata, lasciando agli stessi la possibilità di decidere sul proseguimento del rapporto con l'interlocutore.

Lista risultati parziali attesi Per l'insieme dei risultati attesi a fine progetto, si può fare riferimento alla lista che segue.

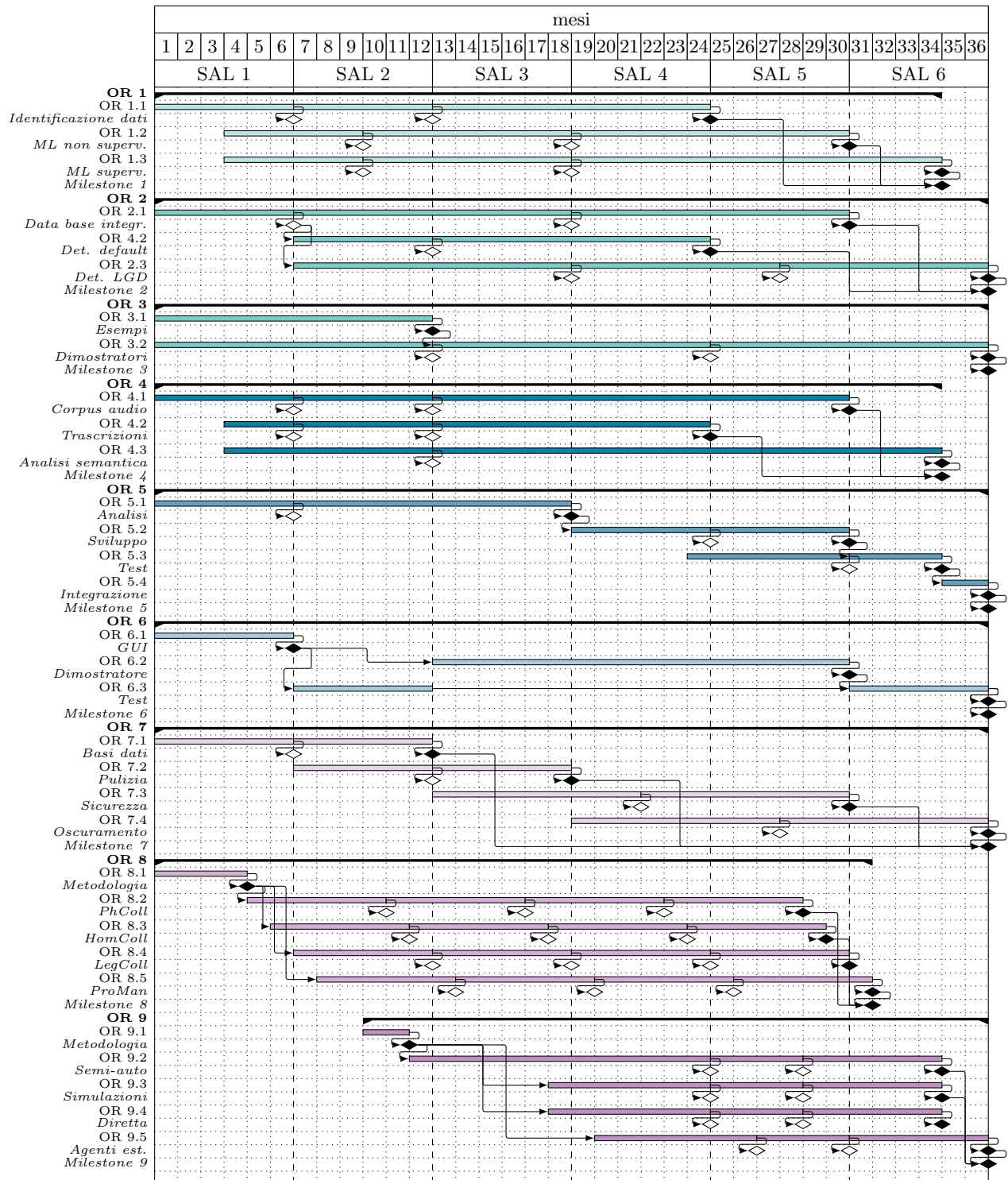
- OR1.1 Identificazione dei dati, etichettatura dei dati e tipologie di analisi
 - 1. Identificazione finale dei dati (strutturati, e dati estratti dalle conversazioni) utili al ML: T0 + 24 mesi
- OR1.2 Algoritmi di ML non supervisionato
 - 1. Retroazione e seconda applicazione degli algoritmi/strumenti ai dati: T0 + 30 mesi
- OR1.3 Algoritmi di ML supervisionato
 - 1. Retroazione e seconda applicazione degli algoritmi/strumenti ai dati: T0 + 34 mesi
- OR2.1 Creazione data base integrato
 - 1. Versione finale: T0 + 30 mesi.
- OR2.2 Principali determinanti del default
 - 1. Serie finale: T0 + 24mesi.
- OR2.3 Principali Determinanti della LGD
 - 1. Serie finale: T0 + 36 mesi.
- OR3.1 Definizione degli esempi per convalida e verifica sperimentale dei dimostratori
 - 1. Identificazione “esempi” per convalida e verifica sperimentale dei dimostratori: T0 + 12 mesi
- OR3.2 Sviluppo e applicazione dei dimostratori
 - 1. Retroazione, e seconda validazione sperimentale utilizzando anche informazioni estratte dalle conversazioni relative agli esempi da OR4 e OR5: T0 + 36 mesi
- OR4.1 Creazione Corpus file audio e documenti taggati
 - 1. Versione finale: T0 + 30 mesi.
- OR4.2 Riconoscimento file audio (ASR)
 - 1. Corpus di test completo (100%): T0 + 24 mesi.
- OR4.3 Analisi linguistica profonda dei testi

1. Versione finale: $T0 + 34$ mesi.
- OR5.1 Analisi del contesto
 1. Output dell'analisi: $T0 + 18$ mesi.
- OR5.2 Sviluppo del sistema
 1. Versione finale: $T0 + 30$ mesi.
- OR5.3 Test sul sistema sviluppato e feedback
 1. Ultimo feedback: $T0 + 34$ mesi.
- OR5.4 Integrazione con gli altri moduli
 1. Sistema integrato in MIRC.0: $T0 + 36$ mesi.
- OR6.1 Progettazione interfaccia grafica
 1. Versione finale: $T0 + 36$ mesi.
- OR6.2 Sviluppo ambiente dimostrativo e di test continuo dell'analisi linguistica
 1. Versione finale: $T0 + 36$ mesi.
- OR6.3 Test con diverse categorie di utenti
 1. Versione finale: $T0 + 36$ mesi.
- OR7.1 Definizione delle basi dati da utilizzare
 1. Elenco basi statistiche o altro e accesso a queste basi: $T0 + 12$ mesi.
- OR7.2 Pulizia dei dati
 1. Sistema di pulizia per basi generiche: $T0 + 18$ mesi.
- OR7.3 Sicurezza dei dati
 1. Sistema funzionante: $T0 + 30$ mesi.
- OR7.4 Oscuramento dei dati
 1. Sistema funzionante: $T0 + 36$ mesi.
- OR8.1 Metodologie di base
 1. Organizzazione dell'OR: $T0 + 4$ mesi;
- OR8.2 Analisi della conoscenza dei phone collector
 1. Output finale: $T0 + 28$ mesi.
- OR8.3 Analisi della conoscenza degli home collector
 1. Output finale: $T0 + 29$ mesi.
- OR8.4 Analisi della conoscenza dei legal collector

MIRC.0

1. Output finale: $T0 + 30$ mesi.
- OR8.5 Analisi della conoscenza dei production manager
 1. Output finale: $T0 + 31$ mesi.
- OR9.1 Metodologie di base
 1. Organizzazione dell'OR: $T0 + 11$ mesi;
- OR9.2 Test semi-automatici
 1. Fine test: $T0 + 34$ mesi;
- OR9.3 Test con simulazione
 1. Fine test: $T0 + 34$ mesi;
- OR9.4 Analisi diretta
 1. Fine test: $T0 + 34$ mesi;
- OR9.5 Interazione con agenti esterni
 1. Fine test: $T0 + 36$ mesi;

10 DIAGRAMMA TEMPORALE DEL PROGETTO



11 ULTERIORI INFORMAZIONI SULLE VOCI DI SPESA PREVISTE NEL PROGETTO

11.1. Fire

Attrezzature

Alla voce Attrezzature si presenta un costo complessivo ammissibile per 20.001,53 €, a fronte di una spesa pari a 62.663,00 €. Il costo ammissibile è stato calcolato in base al periodo di ammortamento dei beni, al periodo di utilizzo nel progetto ed alla percentuale di utilizzo. Si prevede di acquistare di attrezzature informatiche, incrementalmente rispetto alla dotazione aziendale e necessarie per le attività di progetto.

In tabella a seguire la descrizione delle attrezzature previste

N.o	Descrizione	Spesa prevista (€)	Periodo di ammortamento fiscale del bene (mesi)	Periodo di utilizzo nel progetto (mesi)	Percentuale di imputazione al progetto (%)	Costo ammissibile (€)
21	MacBook Apple	52.500,00	60	35	50	15.312,50
1	server	6.200,00	60	35	50	1.808,33
6	stampanti	1.963,00	36	35	100	1.908,47
1	sistema di videoconferenza	2000,00	36	35	50	972,22
		62.663,00				20.001,53

Consulenze

Per le consulenze vengono esposti costi complessivi pari a 438.000,00 €, suddivisi in 214.620,00 € per le attività di RI e 223.380,00 € per le attività di SS. La Fire ha già identificato le strutture cui affidare la consulenza, sulla base delle collaborazioni pregresse e della qualità del fornitore e delle competenze possedute, necessarie alla realizzazione del progetto.

L'Università di Ferrara (le cui competenze sono descritte nella sezione "Qualità delle Collaborazioni") interverrà nella fase di sviluppo e applicazione di soluzioni di Machine Learning (ML) per estrarre conoscenza dalla mole di dati raccolti riguardanti sia le pratiche sia gli interlocutori con l'obiettivo di creare un sistema di classificazione automatica (prodotto attraverso tecniche di ML, partendo dai dati etichettati) che al presentarsi di una nuova pratica (o di un nuovo debitore), sia in grado di restituire una misura prevista del grado di successo nell'affrontare tale pratica. Affiancherà quindi i partner progettuali nelle fasi di identificazione delle variabili (attributi) rilevanti per le varie tipologie di debitore, di pratiche e tipologia di debito e di identificazione dei sistemi di apprendimento automatico (ML) più idonei per identificare similarità e/o per produrre un classificatore automatico da applicare su nuovi casi. La consulenza riguarderà anche lo sviluppo di dimostratori e la loro convalida e verifica sperimentale.

Costo complessivo consulenza: 30.000,00 € di cui 14.700 € RI e 15.300 € SS

La consulenza per il coordinamento tecnico scientifico dei vari soggetti esterni o interni impegnati sul progetto comprende i seguenti aspetti: 1) Organizzazione del progetto (Sistema di guida, Sistema di controllo, Unità operative interne, Consulenti, Interfacce con i partner e con l'esterno del progetto), 2) Piano di lavoro (Verifica del programma, Controllo delle procedure in essere, Controllo delle fonti

interne, Verifica dello stato dell'arte, Definizione del piano); 3) Verifiche in itinere (Controllo dei report intermedi, dei risultati/report finali delle attività, della qualità, Prevenzione delle controversie sui flussi di informazione del progetto, Soluzione delle controversie, Verifica della rispondenza delle disponibilità (uomini/mezzi) con le dotazioni previste); 4) Aggiornamenti (Analisi dei risultati del "technology watching", delle informazioni ottenute ed effetto di queste informazioni sugli obiettivi intermedi/finali, Aggiornamenti del programma di lavoro, della matrice ruoli/responsabilità, delle dotazioni, Adattamento degli obiettivi intermedi, Rimodulazione degli obiettivi finali); 5) Controlli e riconfigurazioni (Analisi della rispondenza azioni/programmazione e della rispondenza funzioni/attribuzioni/risultati, Aggiornamenti del programma di lavoro, della matrice ruoli/responsabilità e delle dotazioni, Adattamento degli obiettivi intermedi, Rimodulazione degli obiettivi finali, Riconfigurazione delle funzioni e/o delle strutture che svolgono le ricerche); 6) Verifiche finali (Controllo dei risultati/report finali delle attività, Controllo di tutte le documentazioni tecniche e scientifiche del progetto, Definizione delle modalità di mantenimento o impiego dei risultati del progetto successivamente alla sua conclusione in modo da conservarne il carattere testimoniale delle attività svolte e dei risultati raggiunti); 7) Sistema di output dei risultati (Output previsti da capitolato, Output per l'operatività del flusso di informazioni interne, Output per l'applicabilità dei risultati al sistema industriale di riferimento (informazione, formazione, supporto delle decisioni, assistenza), Output per la diffusione e la divulgazione).

Il Fornitore individuato è la **GTS Consulting di Messina**, la cui descrizione è fornita al §1.2 della III^a Parte di questo documento.

Materiali e Forniture

Non è previsto l'acquisto di Materiali e Forniture.

11.2. Expert System

Attrezzature

Nessun costo esposto per la voce Attrezzature

Consulenze

Per le Consulenze sono esposti costi per 200.000 €, di cui 98.000 € per le attività RI e 102.000 € per le attività di Sviluppo.

Si prevede di attivare consulenza in ambito di **riconoscimento vocale (ASR)** - Modelli acustici e trattamento automatico di contenuti multimediali. In questo caso il fornitore non è stato ancora individuato, e il costo è stato valutato tendo conto dei valori di mercato nel settore altamente specialistico.

Costo complessivo consulenza: **200.000,00 €, di cui 98.000 € RI e 102.000 € SS.**

Expert System per lo sviluppo delle attività ha inoltre necessità di acquistare ed utilizzare esclusivamente per il progetto la licenza per motore di riconoscimento vocale. La tipologia di motore è individuato sul mercato con le caratteristiche più adatte al contesto di riferimento

Al momento i software di riconoscimento vocale leader del mercato sono infatti tutti di tipo proprietario, e per questo non completamente sfruttabili dagli sviluppatori.

Il corrispondente costo, pari ad 50.000 €, di cui 24.500 € RI e 25.500 € SS, viene esposto alla voce Beni Immateriali.

Tale bene sarà utilizzato esclusivamente per il progetto, e per tale motivo è stato esposto il costo di acquisto.

Expert System svolge attività anche in Emilia Romagna (Regioni più sviluppate), ma esponendo solo costi del personale e relative spese generali. I costi precedentemente descritti afferiscono totalmente alla sede di Napoli (Regioni meno sviluppate).

Materiali e Forniture

Non è previsto l'acquisto di Materiali e Forniture.

11.3. Università di Messina

Alla voce Attrezzature si presenta un costo complessivo ammissibile per 5.625,28 €, a fronte di una spesa pari a 15.641,00 €. Il costo ammissibile è stato calcolato in base al periodo di ammortamento dei beni, al periodo di utilizzo nel progetto ed alla percentuale di utilizzo. Si prevede di acquistare di attrezzature informatiche, incrementali rispetto alla dotazione attuale dell'Università e necessarie per le attività di progetto.

In tabella a seguire la descrizione delle attrezzature previste

N.o	Descrizione	Spesa prevista (€)	Periodo di ammortamento fiscale del bene (mesi)	Periodo di utilizzo nel progetto (mesi)	Percentuale di imputazione al progetto (%)	Costo ammissibile (€)
4	MacBook Apple	10.000,00	60	35	50	2.916,67
1	server	2.650,00	60	35	50	772,92
3	stampanti	991,00	36	35	100	963,47
1	sistema di videoconferenza	2000,00	36	35	50	972,22
		15.641,00				5.6285,28

Consulenze

Per le consulenze vengono esposti costi complessivi par a 218.000,00 €, suddivisi in 115.712,00 € per le attività di RI e 103.088,00 € per le attività di SS.

Unime ha identificato nell'Università di Ferrara, sulla base delle collaborazioni pregresse e della qualità del fornitore e delle competenze possedute, il fornitore per lo sviluppo e l'applicazione di soluzioni di Machine Learning (ML) per estrarre conoscenza dalla mole di dati raccolti riguardanti sia le pratiche sia gli interlocutori con l'obiettivo di creare un sistema di classificazione automatica (prodotto attraverso tecniche di ML, partendo dai dati etichettati) che al presentarsi di una nuova pratica (o di un nuovo debitore), sia in grado di restituire una misura prevista del grado di successo nell'affrontare tale pratica. Affiancherà quindi i partner progettuali nelle fasi di identificazione delle variabili (attributi) rilevanti per le varie tipologie di debitore, di pratiche e tipologia di debito e di identificazione dei sistemi di apprendimento automatico (ML) più idonei per identificare similarità e/o per produrre un classificatore automatico da applicare su nuovi casi. La consulenza riguarderà anche lo sviluppo di dimostratori e la loro convalida e verifica sperimentale.

Per la seconda consulenza prevista, il cui importo è stato valutato, sulla base dell'impegno previsto e dei normali valori di mercato, in complessivi 48.800 €, di cui 23.912 € RI e 24.888 € SS.

La consulenza prevederà attività per l'organizzazione del progetto, la verifica del Piano di lavoro attraverso verifiche in itinere, aggiornamenti, controlli e riconfigurazioni e verifiche finali per ogni OR e complessive, oltre al coordinamento con gli altri soggetti del partenariato.

Il fornitore verrà selezionato con gara pubblica.

Materiali e Forniture

Non è previsto l'acquisto di Materiali e Forniture.

11.4. Spese fuori area Obiettivo

Il progetto prevede lo svolgimento di attività anche in aree del territorio nazionale non comprese nelle regioni meno sviluppate e nelle regioni in transizione. Di seguito sono riportati i costi di progetto costo sostenute al di fuori di dette regioni e gli elementi utili a comprovare il rispetto della soglia massima del 35% prevista.

MIRC.0

SOGGETTI	COSTI TOTALI (€)	COSTO ALTRE AREE (€)	INCIDENZA (%)
FIRE	2.833.496,90	0,00	
EXPERT SYSTEM	1.251.625,00	657.112,50	
UNIME	903.721,60	0,00	
TOTALE	4.988.843,50	657.112,50	13,17

12 ELEMENTI A SUPPORTO DELLA RICHIESTA DI MAGGIORAZIONE DEL CONTRIBUTO

La strutturazione del progetto soddisfa le condizioni previste dal DD per il riconoscimento della maggiorazione del contributo. Si tratta infatti di un progetto congiunto, in cui:

- nessuno dei partecipanti privati sostiene da solo più del 70% dei costi di progetto
- sono tra loro indipendenti.

Si configura quindi la condizione di “collaborazione effettiva” come definita dal DD.

Inoltre, entrambi i partner privati appartengono alla categoria delle PMI (nello specifico Fire ed Expert System sono entrambe medie imprese) soddisfacendo quindi le condizioni richieste per la maggiorazione del contributo (collaborazione effettiva tra imprese di cui almeno una appartenente alla categoria dell PMI).

Di seguito la ripartizione dei costi di progetto atta a confermare le percentuali di costo per singolo soggetto

Soggetto	Costi ammissibili esposti in €	% su totale costi
FIRE	2.833.496,90	56,80
EXPERT SYSTEM	1.251.625,00	25,09
UNIME	903.721,60	18,11
Totale	4.988.843,50	100,00

III^a PARTE: ELEMENTI VALUTATIVI

1 CARATTERISTICHE DEL SOGGETTO PROPONENTE

1.1. CAPACITÀ TECNICO-ORGANIZZATIVA

Fire Spa

La Fire Spa, è perfettamente in grado di realizzare il progetto con risorse interne. In particolare la struttura dedicata alle attività di R&D si avvale della collaborazione dei reparti Process Governance, Innovation ed Analytics composte in totale da 12 figure professionali la cui formazione comprende le seguenti aree di competenza:

- Ingegneria Gestionale
- Statistica
- Matematica
- Informatica
- Economia
- Lean Organization
- Six Sigma
- Impact Innovation

La Società nel 2016 ha portato a termine un progetto di Ricerca avente ad oggetto la creazione di un processo di recupero innovativo, misurabile e più efficace ed efficiente delle value stream. Sul fronte dell'innovazione, il progetto prevedeva una fase di breaktrough innovation le cui linee guida possono essere così sintetizzate:

- definizione di nuove idee per l'innovazione del processo di recupero crediti in linea con le strategie aziendali (nuovi processi, nuovi strumenti, nuove tecnologie, etc.),
- sperimentazione nuovi metodi e strumenti in grado di far ottenere risultati in tempi brevi sul piano dell'innovazione sia tecnologica che di processo.

L'importo sostenuto per il progetto ammonta a 131.000,00 € nel 2016. Il progetto è stato cofinanziato con il credito d'imposta.

Inoltre, nel 2018, ha iniziato un progetto di ricerca destinato a individuare quali variabili – che riguardano tanto caratteristiche relative alla pratica stessa, quanto invece relative al debitore, quanto infine allo stato generale della situazione economica del Paese – influenzano una misura di redditività della soddisfazione della pratica stessa. L'importo di questo progetto, che sarà cofinanziato con il credito d'imposta, è di circa 720.000,00 € (conteggi in corso).

Per l'attività di ricerca e sviluppo del presente progetto verranno utilizzate le infrastrutture informatiche già in uso in azienda, ed in particolare:

- piattaforma di sviluppo Microsoft ASP/Net e base dati su Microsoft SQL Server, disponibilità di tecnologie di data encryption trasparente basata su standard di mercato;

- architettura server con S.O. Microsoft, virtualizzata su tecnologia VMWare, per consentire rapidi scale-up e scale-out delle risorse;
- database ospitati su infrastruttura MS-SQL Server clusterizzata per garantire alta affidabilità ed alta disponibilità;
- sito di Disaster Recovery remoto
- Tools di produttività, Data mining e reporting, Business Intelligence e Artificial Intelligence: Microsoft Office Suite; Microsoft Visio, Microsoft Machine Learning; SAP Business Object, Microsoft Power BI.

Expert System Spa

Expert System spa Unica azienda italiana e fra le pochissime al mondo ad essere contemporaneamente presente sia nel Magic Quadrant di Gartner dedicato agli Insight Engine sia nel report di Forrester dedicato alle text analytics Big Data Platform, Expert System ha di recente ottenuto la qualifica di Pmi Innovativa dalla Camera di Commercio di Trento, qualifica che riconosce l'ampia propensione della società verso l'innovazione tecnologica nell'evoluzione del proprio modello di business. Expert System sostiene numerosi progetti di ricerca nel campo dell'intelligenza artificiale per promuovere i progressi tecnologici e il processo di trasformazione digitale nell'ambito della pubblica amministrazione e dei diversi settori industriali. La collaborazione fra organizzazioni, enti pubblici, dipartimenti universitari e aziende all'avanguardia favorisce infatti l'attivazione di importanti sinergie tra competenze e conoscenze diverse, per realizzare soluzioni tecnologiche innovative e nuovi modelli di business a supporto della crescita economica.

Partecipando attivamente a progetti di ricerca e con la presenza dei "Cogito Lab" in Europa e Stati Uniti, Expert System conferma di puntare sull'innovazione tecnologica come asset strategico fondamentale per lo sviluppo del business. 35 gli addetti che si occupano di Ricerca e Sviluppo, e che attualmente lavorano sui progetti di ricerca a cui l'azienda partecipa. Tra questi citiamo:

Courier – L'Europa sta affrontando crescenti minacce terroristiche da parte di organizzazioni di diversa natura. Un tema rilevante è la propaganda di queste organizzazioni ed il processo di radicalizzazione. Il canale più usato è internet "surface" e deep/dark con modalità sempre più basate su l'utilizzo integrato di testo, video ed immagini. Le agenzie di Intelligence e le forze di Polizia hanno necessità di dotarsi di strumenti capaci di elaborare in modo efficace questi dati, al fine di emanare le direttive operative in tempi rapidi e con elevato grado di attendibilità. In questo contesto si inquadra COURIER che ha l'obiettivo di supportare il decisore con una piattaforma di analisi e con gli elementi rilevanti (frasi, persone, organizzazioni, luoghi, relazioni). COURIER consentirà una risposta tempestiva e mirata delle Agenzie/Dipartimenti di Polizia alle minacce crescenti e sempre più nascoste in un "mare" di dati, facendo emergere "early warnings". Le modalità di visualizzazione saranno su mappa geografica, al fine di consentire una geo-referenziazione degli eventi estratti dai contenuti.

Durata progetto: 18 mesi – Giugno 2018 – Dicembre 2020

Tipo di Finanziamento



Network GCA è un progetto che mira alla realizzazione di un modello innovativo di assistenza sanitaria e sociale, basato su tecnologie dell'informazione e della comunicazione, finalizzato a sostenere la diagnosi, il trattamento e la gestione nei pazienti con Grave Cerebrolesione Acquisita in fase acuta e post-acuta attraverso la realizzazione di un sistema di sanità elettronica e l'integrazione di sistemi di diagnosi per immagini e tecnologie per la riabilitazione e l'assistenza.

Verrà sviluppato un sistema che consenta di seguire il paziente in tutte le fasi, aumentando l'appropriatezza della cura e diminuendo l'impegno economico per la Regione, il tutto attraverso l'integrazione del percorso assistenziale dalla fase iniziale acuta presso le strutture ospedaliere, dove sono stati trattati

in unità operative ad alta intensità di cura, alla fase di riabilitazione cognitivo-comportamentale di pazienti post-coma sino alla dimissione protetta, con l'attivazione della rete di strutture sanitarie regionali chiamate a garantire al paziente la disponibilità di macchinari e ausili in caso di ritorno a casa oppure l'accoglienza in CRE e RSA.

Durata progetto: 18 mesi – Aprile 2018 – Dicembre 2020

Tipo di finanziamento:



Anita progetterà e svilupperà un nuovo sistema di investigazione centrato sull'utente e basato sulla conoscenza per analizzare eterogenei (testo, audio, video, immagini) online (Surface Web, Deep Web, DarkNet) e contenuti offline per combattere il traffico illegale di droghe, farmaci contraffatti, NPS e armi da fuoco. ANITA combinerà a) un'innovativa analisi delle fonti di dati della rete criptovaluta e delle transazioni e delle tecnologie blockchain; b) strumenti avanzati di analisi dei Big Data per l'analisi automatica delle vaste quantità di contenuti multimodali delle fonti identificate; c) metodologie sofisticate per acquisire, modellare e inferire le conoscenze in forme comprensibili dall'uomo (ad esempio ontologie espressive), estraendo anche conoscenze pertinenti e nuove dalle reti neurali e archiviandole formalmente sotto forma di ontologie; d) sviluppo di una struttura di modellizzazione adattativa e cognitiva dell'utente che catturerà, analizzerà, interpreterà, imiterà e integrerà le principali funzioni cognitive e di elaborazione delle informazioni umane; e) applicazioni di intelligenza orientate al dominio e orientate all'utente, che consentiranno agli utenti di identificare i modelli per correlazioni spaziali, temporali e causali tra eventi, entità e attività di traffico illecito e di supportare i processi decisionali per le contromisure da intraprendere.

Durata progetto: 36 mesi - Maggio 2018 – Aprile 2021

Tipo di finanziamento: H2020- RIA – SEC-12-FCT-2016-2017 – Project ID: 787061



Trivalent . L'ipotesi alla base della proposta di progetto, in linea con le raccomandazioni del Consiglio di sicurezza delle Nazioni Unite e la Commissione "Agenda europea sulla sicurezza" 2015-2020, è che per contrastare con successo l'estremismo violento, è necessaria una risposta più equilibrata al terrorismo, combinando misure repressive (protettive) con misure preventive, in un approccio globale in collaborazione con gli attori della società civile e le comunità di riferimento, basato su un fermo impegno a rispettare i diritti fondamentali, promuovendo l'integrazione, il dialogo culturale e combattendo la discriminazione. A tal fine è necessaria una migliore comprensione dei fattori che costituiscono la radicalizzazione violenta

MIRC.0

in Europa, che mira, attraverso un'analisi multidisciplinare, a una visione complessiva del fenomeno, indagando le sue cause profonde, al fine di sviluppare contromisure adeguate.

Durata del progetto: 36 mesi – Maggio 2017 – Aprile 2020

Tipo di finanziamento: sviluppo di un approccio globale alla radicalizzazione violenta nell'UE dalla comprensione precoce al miglioramento della protezione – H2020-SEC-2016-2017-1 RIA –

Azione per la ricerca e l'innovazione – ID progetto: 740934



Di seguito alcuni dei più significativi progetti chiusi nel 2018:

CREEP Lo scopo di CREEP è quello di fornire una serie di strumenti per supportare l'individuazione e la prevenzione dei problemi psicologici / comportamentali delle vittime di cyberbullismo. L'obiettivo viene raggiunto combinando il monitoraggio dei social media e le tecnologie motivazionali (coach virtuali che integrano i chatbot). Il cyberbullismo può innescare depressione, uso di sostanze e tentativi di suicidio di giovani vittime. CREEP fornirà alle parti interessate (ad es. Scuole, autorità locali / nazionali, fornitori di servizi Internet) gli strumenti su misura per prevenire e affrontare tali fenomeni. Il progetto aprirà una nuova branca nel mercato della salute preventiva e personalizzata, che è uno dei settori del benessere che vede la crescita più significativa negli ultimi anni.

Durata progetto: 12 mesi – Gennaio 2018 – Dicembre 2018

Tipo di finanziamento:



Pasteur sviluppa una piattaforma Intelligent Intelligence Artificiale basata sull'Intelligenza Semantica per la consultazione estesa e multidimensionale di contenuti medico-sanitari, estratti da una varietà di fonti testuali di diverse tipologie. È articolato nei seguenti tre moduli software: Semantic Engine, Knowledge Base Normalizer e Knowledge Map Navigator.

Durata progetto: 23 mesi – Febbraio 2017 – Dicembre 2018

Tipo di finanziamento:





VIS sviluppa e distribuisce agli editori il primo sistema completo per il fact-checking assistito direttamente integrato nelle piattaforme esistenti di gestione dei contenuti, ma accessibile anche a individui con un impatto rilevante su tutto l'ecosistema di produzione di notizie. Basandosi anche su soluzioni già disponibili sul mercato, VIS sviluppa un prodotto pronto per le notizie. Il cuore del sistema è la sua funzionalità Supersearch, che utilizza le tecnologie di elaborazione del linguaggio naturale e il motore Cogito di Expert System per cercare una varietà di fonti, sia aperte che proprietarie, combinate con l'analisi dei social network.

Durata progetto: 18 mesi - Settembre 2017 - Dicembre 2018

Tipo di finanziamento: Google DNI – Round 3



Mixed emotion è una Linked Data platform per l'analisi delle emozioni raccolte dall'analisi dei contenuti scritti dagli utenti nei social network a commento dei programmi TV che stanno guardando. Mixed Emotions analizza il profilo emozionale del comportamento del pubblico televisivo utilizzando i dati provenienti da fonti differenti: testi multilingue, audio, video, social media e dati strutturati. Le applicazioni commerciali (implementate come progetti pilota) saranno rivolte inizialmente a Social TV, alla gestione della Brand Reputation ed a funzioni evolute per gli operatori di Call Centres.

Durata progetto: 24 mesi – Febbraio 2015 – Gennaio 2017

Tipo di finanziamento: EU Horizon 2020 ICT-15-2014 Big data and Open Data Innovation Funding Body – Industrial Leadership (IL) – Number of Project H2020 141111



Nel 2017 i costi capitalizzati per Ricerca e sviluppo ammontano ad 3.225.947 €, a fronte di un fatturato di 19.254.067 €.

Università di Messina – Dipartimento di Economia

L'Università di Messina partecipa al progetto come soggetto co-proponente con il Dipartimento di Economia.

MIRC.0

Il Dipartimento di Economia dell'Università di Messina nell'ultimo triennio è stato impegnato attivamente nella realizzazione del seguente progetto di ricerca coerente con gli obiettivi del presente bando:

- "Higher moments of asset returns: realized measures, the cross-section of equity returns and mutual fund returns, volatility clustering and correlation determinants" programma Research & Mobility (responsabile scientifico prof. Walter Distaso) 79.735,00 €.

Fra le recenti ricerche che hanno goduto di finanziamenti esterni si segnalano:

- Progetto di ricerca "R.A.M. Project, Rethinking Anchor Model, Assessing Local Sustainability through Firm Strategies and Networks" commissionato da Raffineria di Milazzo (responsabile scientifico prof.ssa Daniela Baglieri);
- "A novel approach to Urban Metabolism: integration of economic, environmental and social issues for the design of sustainable urban systems" programma Research & Mobility (responsabile scientifico prof. Giuseppe Ioppolo) 72.750,00 €.

Il personale docente del Dipartimento di Economia ha inoltre svolto attività di ricerca sui profili economico-finanziari dei seguenti progetti incardinati presso altre strutture dell'Ateneo di Messina:

- PANLAB (PON Research and Competitiveness 2007-2013) "Progetto di potenziamento strutturale di laboratori dell'Università di Messina per analisi degli alimenti, studio della loro incidenza sulla salute umana e consulenza tecnologica, giuridica ed economica alle aziende agroalimentari".? PON PANLAB A3_00166, CUP J41D11000080007 (22.500.000,00 €);
- Be&Save (PON Research and Competitiveness 2007-2013) Tecnologie e modelli operativi per la gestione sostenibile della filiera alimentare attraverso la valorizzazione degli scarti biologici della produzione a scopi energetici, la riduzione degli sprechi alimentari del sistema distributivo e dei consumatori e il trattamento e la valorizzazione della frazione edibile del rifiuto solido urbano"- Smart Cities and Communities Prot. 84/Ric del 02/03/2012, Asse II – Sostegno all'innovazione, CUP J41H12000030005 (4.722.067,53 €).

Nello stesso arco temporale del progetto proposto saranno realizzati anche i seguenti progetti:

- "Goforit: L'esperienza prima della partenza" - Programma Operativo Nazionale "Ricerca e Innovazione" (PON "R&I" 2014-2020) – Responsabile scientifico prof.ssa Roberta Salomone CUP J66C18000370005 (€3.849.999,99).
- "Cresting: circular economy sustainability Implications and guiding process" – Programma Horizon 2020 Marie Skłodowska – Responsabile scientifico prof.ssa Roberta Salomone CUP J43C17000390006 (€235.810,81);
- "I Know: trasferimento tecnologico, diffusione della cultura d'impresa, miglioramento della qualità della vita" Programma V Interreg Italia-Malta – Responsabile scientifico prof. Andrea Cirà CUP J51G18000000007 (€228.954,25)

Il Responsabile Scientifico del partner UNIME è il professore Walter Distaso di cui si elencano alcune qualifiche e indicazioni in merito alle capacità di coordinamento: nato a Bari il 12/10/1969. Laurea in Economia e Commercio 110 e lode, UNIBA (1991). PhD in "Statistica" (1997), UNIBA. PhD in "Economics" (2002), University of York. Full Professor of "Financial Econometrics", Imperial College Business School, Imperial College London (2010). Professore Ordinario di "Econometria", Dip. Economia, Università degli Studi di Messina (2010).

Responsabile di progetti di ricerca italiani ed europei tra cui:

- 2012-2015 proposta finanziata da NERC (£2.3m) su "Will climate change in the Arctic increase the landslide-tsunami risk to the UK?".
- 2011-2012 British Academy Mid-Career Fellowship su "Risk management implications of weather-related extreme events".

- 2009-2016 PI per la KIC su climate change, con un cash flow di £75 milioni.
- 2010-2011 Ricerca finanziata (£40k) da EPSRC strategic fund grant, su “Modeling and Risk Management Implications of Weather-related Extreme Events”, con Sir Brian Hoskins e Enrico Biffis.
- 2007-2010 Ricerca finanziata (£390k) da ESRC grant, su “Extraction of trend, cycle, and memory from economic and financial series”, con Karim Abadir e Liudas Giraitis.
- 2007-2009 Ricerca finanziata (£385k) da ESRC grant (PI), su “Conditional Independence, Noncausality and International Market Links: A Realized Measure Approach”, con Valentina Corradi e Marcelo Fernandes.
- 2002-2005 Ricerca finanziata (£140k) da ESRC grant R000230006, su “Modeling, testing and prediction of integrated volatility through realized volatility”, con Valentina Corradi.
- 2001-2004 Ricerca finanziata (£120k) da ESRC grant R000239538, su “Improved inference for partially one-sided hypotheses”, con Karim Abadir.
- 2015 responsabile scientifico del progetto "Higher moments of asset returns: realized measures, the cross-section of equity returns and mutual fund returns, volatility clustering and correlation determinants" programma Research & Mobility, €79.735,00.

Altre attività inerenti l'ambito del progetto:

- Direttore Scientifico del “FI.RE. PHA (Fire Centre for Research on Payment Habits Analysis)”, che persegue la finalità di promozione e svolgimento di attività di formazione e di ricerca, sia teorica che empirica, sui temi della finanza, con particolare attenzione alla gestione del risparmio e al recupero del credito;
- Referee per proposte di ricerca presso ESRC, UK, EPSRC, UK, Open Programme, Netherlands Organization for Scientific Research.
- Research Director, FVC llp.
- Componente del gruppo di esperti di valutazione (SECSP05) per la valutazione della qualità della ricerca (VQR) (2016).
- Ha pubblicato più di 20 articoli scientifici su riviste internazionali.?

Le attività realizzate dal Dipartimento di Economia dell'Università degli Studi di Messina, per l'esecuzione del progetto presentato nell'ambito del bando MISE PON “Imprese e Competitività” 2014 – 2020 - Fondo crescita Fabbrica Intelligente, Agrifood Procedura a sportello DD 20 novembre 2018, si svolgeranno nei locali e nei laboratori siti nel plesso centrale dell'Ateneo ed assegnati al Dipartimento. In particolare, sarà messo a disposizione un laboratorio informatico dotato di trenta postazioni provvisto di hardware e software di ultimissima generazione in grado di svolgere elaborazioni economico-finanziarie particolarmente sofisticate.

Inoltre è in fase di trasferimento dai locali di viale Italia (sempre del Dipartimento di Economia) ove era ubicato al plesso centrale l'Experimental Lab. Anch'esso dotato di 30 postazioni in grado di fornire un adeguato supporto all'attività di ricerca in campo economico e statistico. Il laboratorio costituisce, inoltre, un luogo di coordinamento e discussione per coloro che utilizzano esperimenti economici-statistici per ricerca ed in particolare per quel ramo dell'economia denominata “sperimentale”.

1.2. QUALITÀ DELLE COLLABORAZIONI

Uno dei tre partner coinvolti nel progetto è l'Università di Messina, che – oltre a mettere in gioco le proprie strutture di ricerca descritte nei capitoli precedenti – si avvarrà della collaborazione dell'Università di Ferrara, in particolare del Dipartimento di Ingegneria e del Laboratorio MechLav del Tecnopolo di questa Università.

Il progetto è stato concepito e la proposta progettuale è stata coordinata e redatta dalla GTS Consulting, che parteciperà col compito di organizzare e coordinare le attività.

Descrizione del Dipartimento di Ingegneria dell'Università di Ferrara

Il Dipartimento di Ingegneria dell'Università di Ferrara è nato nel 1996. Nel 2012, in seguito alla Legge 240/10, si è costituito come nuova struttura che svolge attività di formazione e ricerca negli ambiti dell'Ingegneria Civile, Industriale e dell'Informazione. Al Dipartimento afferiscono oltre ottanta unità di personale di ruolo fra docenti, ricercatori, personale tecnico e amministrativo e circa un centinaio tra assegnisti di ricerca, dottorandi e collaboratori alla ricerca che operano per offrire il meglio nella ricerca e nella didattica, con corsi di Laurea in Ingegneria Civile e Ambientale, Ingegneria Elettronica e Informatica, Ingegneria Meccanica e Lauree Magistrali in Ingegneria Civile, Ingegneria Elettronica e delle Telecomunicazioni, Ingegneria Informatica e dell'Automazione, Ingegneria Meccanica.

Il Dipartimento di Ingegneria dell'Università di Ferrara è organizzato in gruppi di ricerca e, per attività specifiche, si struttura in maniera più articolata. In particolare, fanno fondamentalmente capo al Dipartimento, il Centro di Studi sulla Corrosione e la Metallurgia Aldo Daccò e il Laboratorio MechLav del Tecnopolo dell'Università di Ferrara, mentre vari docenti di Dipartimento collaborano con altri Laboratori del Tecnopolo operanti nell'Università di Ferrara: Terra&AcquaTech e Teknehub.

Le principali linee di ricerca sviluppate dai gruppi di ricerca riguardano:

- **INGEGNERIA CIVILE:** Acustica; Architettura tecnica e produzione edilizia; Idraulica, Idrologia, Costruzioni idrauliche; Economia agro-ambientale ed Estimo; Geotecnica e fondazioni; Ingegneria ambientale e sanitaria; Ingegneria strutturale; Scienza delle costruzioni; Topografia e cartografia;
- **INGEGNERIA MECCANICA:** Costruzioni di macchine; Fisica tecnica industriale; Macchine a fluido e Sistemi energetici; Materiali polimerici e compositi; Meccanica delle macchine e vibrazioni; Metallurgia e corrosione; Tecnologia meccanica;
- **INGEGNERIA DELL'INFORMAZIONE:** Automazione; Campi elettromagnetici ed ottica; Elaborazione dei segnali; Elettronica digitale ed affidabilità; Elettronica delle telecomunicazioni; Intelligenza artificiale; Ricerca operativa; Sistemi distribuiti; Telecomunicazioni.

Le attività di ricerca sono sviluppate sia nell'ambito di finanziamenti pubblici sia in collaborazione con imprese private nazionali ed estere.

Dall'ultimo monitoraggio effettuato (Luglio 2017) risultano attivati, nel solo triennio 2015-2017, 39 progetti di ricerca istituzionale, e 6 ulteriori progetti europei risultano attivi nel 2017 (uno del VII PQ, 5 di H2020). Sono attivi nel solo 2017 progetti di ricerca finanziati da enti pubblici per un valore complessivo di oltre 4,1 M€.

Le collaborazioni e contratti di ricerca con Aziende sono numerosissimi e rappresentano una delle attività principali del Dipartimento con contratti stipulati per oltre 1,4 M€ nel triennio 2015-2017. In questo ambito sono spesso affrontati argomenti di particolare rilevanza scientifica o situazioni in cui le usuali regole di progettazione sembrano non fornire più risultati adeguati. Inoltre, nello stesso contesto trovano spazio anche tutte quelle prove che richiedono l'utilizzo della strumentazione scientifica avanzata di cui dispone il Dipartimento. Per la ricerca applicata, commissionata e il trasferimento tecnologico, nel triennio 2015-2017, il Dipartimento ha siglato 64 contratti di ricerca commissionata, di cui 16 stipulati per i Laboratori del Tecnopolo, per un valore medio annuo di contratti di circa 1,5 M€.

MechLav è un laboratorio del Tecnopolo dell'Università di Ferrara ed è struttura accreditata di ricerca industriale della Rete Alta Tecnologia dell'Emilia-Romagna, socio fondatore di 4 Clust-ER regionali e del Cluster Tecnologico Nazionale Fabbrica Intelligente. All'interno di MechLav, l'Università dispone mediamente di 18 unità di personale dedicato a tempo pieno alle attività di ricerca industriale, sviluppo sperimentale e trasferimento tecnologico e di 31 unità di personale universitario strutturato.

MechLav svolge un'intensa attività di ricerca industriale e trasferimento tecnologico verso le Imprese, conducendo Progetti di Ricerca Commissionata per un ammontare medio annuo di circa 430 k€. Partecipa a molti progetti competitivi di ricerca industriale a livello nazionale e regionale, nell'ambito del Cluster Tecnologico Nazionale Fabbrica Intelligente e nell'ambito Smart Cities Nazionali (tra cui il progetto SORT - Sviluppo di sistemi tecnologici innovativi integrati per lo Spacchettamento, l'Organizzazione delle scorte e il Tracciamento dei prodotti alimentari sprecati finalizzati alla loro valorizzazione). Inoltre MechLav è stato partner di due progetti finanziati dal Bando Emilia-Romagna DGR 774-201 (tra cui il progetto GST4Water - Green-Smart Technology per l'utilizzo sostenibile della risorsa idrica negli edifici e in ambito urbano).

MechLav ha competenze in molti ambiti: tecnologie ICT per l'Industria 4.0 (in particolare Big Data analytics, Machine learning, Cloud Computing, IoT), automazione industriale, meccanica, termofluidodinamica, vibro-acustica, manutenzione avanzata.

Il gruppo di ricerca di intelligenza artificiale del Dipartimento di Ingegneria e del laboratorio MechLav (Università di Ferrara) svolge attività di ricerca scientifica in diversi ambiti, quali quello dei sistemi knowledge-based, dei sistemi di apprendimento automatico, dei sistemi a vincoli, e del riconoscimento di immagini.

Su queste tematiche i membri del gruppo svolgono ricerca di base, proponendo, realizzando e sperimentando nuovi algoritmi e sistemi di ragionamento, apprendimento automatico e che usano propagazione di vincoli per cercare soluzioni in problemi vincolati. Il gruppo è riconosciuto come leader a livello internazionale per le sue competenze su sistemi di ragionamento e apprendimento probabilistico.

Le tecniche di intelligenza artificiale - di cui il gruppo è altamente competente, con esperienza pluridecennale - risultano oggi fondamentali anche in ambito industriale, poiché nella visione di Industry 4.0, la quarta rivoluzione industriale è guidata dall'introduzione nella fabbrica di tecnologie ICT avanzate, tra cui quelle basate su intelligenza artificiale. La ricerca applicata svolta dal gruppo ha sempre avuto l'obiettivo di calare queste tecnologie altamente strategiche in contesti applicativi concreti, quali quello sanitario, industriale, etc.

Il gruppo ha partecipato allo sviluppo, realizzazione e sperimentazione di un sistema di supporto alle decisioni knowledge-based per migliorare l'appropriatezza degli esami di laboratorio in ambito clinico-medico, in collaborazione con l'Azienda Universitaria Ospedaliera di Ferrara, e l'azienda Noemalife S.p.a., finanziato da Ministero della Salute – Ricerca Finalizzata 2010 (Progetto “Decision support system to improve appropriateness of repeated execution of laboratory examinations”). Obiettivo del progetto è stato quello di realizzare un sistema di supporto alle decisioni, basato su regole, che – in fase di prescrizione di esami di laboratorio a scopi diagnostici - propone al medico di reparto ospedaliero di evitare la prescrizione di alcuni di essi ove siano già disponibili nel sistema informativo di laboratorio esiti non obsoleti o non critici.

I membri del gruppo hanno partecipato inoltre a progetti nazionali e regionali, tra cui i più recenti: (i) il progetto nazionale SORT (“Sviluppo di sistemi tecnologici innovativi integrati per lo Spacchettamento, l'Organizzazione delle scorte e il Tracciamento dei prodotti alimentari sprecati finalizzati alla loro valorizzazione” – Development of innovative integrated technological systems for the unpacking, stocks and tracking of wasted food products for their valorisation) finanziato da MIUR “Programma Operativo Nazionale (PON) Ricerca e Competitività 2007-2013”, Smart Cities and Communities and Social Innovation; (ii) il progetto regionale “Green Smart Technology for water (GST4Water) per l'utilizzo sostenibile della risorsa idrica negli edifici e in ambito urbano” finanziato da bando POR-FESR 2014-2020; (iii) il progetto PON2007-2013 CittàEducatrice, come sub-contractor, per l'applicazione di tecniche di apprendimento e ragionamento probabilistico a dati estratti da video.

Obiettivo del progetto SORT è ridurre lo spreco di prodotti alimentari permettendone il riciclo intelligente attraverso l'estesa adozione di tecnologie ICT avanzate. All'interno del progetto SORT, il gruppo di ricerca di intelligenza artificiale si è occupato del riconoscimento automatico delle date di scadenza, e del numero di lotto dei prodotti, e ha concorso alla realizzazione del sistema di supporto alle decisioni che deve ottimizzare il recupero dei prodotti.

L'obiettivo del progetto GST4Water è stato quello di sperimentare alcune soluzioni innovative, orientate al risparmio e alla sostenibilità, nel settore idrico, attraverso l'introduzione delle nuove tecnologie in ambito ICT, per rendere i cittadini più consapevoli e attenti ai propri comportamenti di consumo dell'acqua. All'interno del progetto GST4Water, il gruppo di ricerca di intelligenza artificiale si è occupato della realizzazione di un modulo classificatore, appreso applicando tecniche di machine learning a dati disponibili sui consumi di oltre 200 famiglie raccolti nell'arco di un anno ed etichettati, per classificare e identificare quindi precocemente perdite idriche.

Il progetto CittàEducatrice intende contribuire a generare una "città educatrice", ossia un modello in cui il momento educativo è basato sulla reciprocità dei soggetti coinvolti. Nel progetto, si sono estratte caratteristiche da video a contenuto culturale, e il gruppo di intelligenza artificiale dell'Università di Ferrara si è occupato dell'applicazione di tecniche di apprendimento probabilistico ai dati estratti, per costruire una ontologia di concetti, su cui effettuare anche ragionamento probabilistico, utile all'indicizzazione e alla ricerca dei video.

Le competenze sviluppate durante i summenzionati progetti CittàEducatrice, GST4Water e SORT potranno certamente essere messe a frutto nel progetto di ricerca illustrato nella presente proposta.

La coordinatrice del gruppo di ricerca su Intelligenza Artificiale presso è Evelina Lamma, Professore di I fascia presso il Dipartimento di Ingegneria dell'Università di Ferrara. Presso tale Università, è stata coordinatore del Dottorato in Scienze dell'Ingegneria per un triennio (2004-2007) e coordinatore di tre corsi di studio per 8 anni (2007-2015). E' stata Delegata alla Didattica (area scientifico-tecnologica) per il triennio 2015-2018. Dal 2017 è Presidente del Presidio Qualità di Ateneo. E' delegato di Ateneo nel Consiglio Direttivo del CINI - Consorzio Interuniversitario Nazionale per l'Informatica.

La sua attività di ricerca ha riguardato e riguarda l'intelligenza artificiale e lo sviluppo di sistemi di supporto alle decisioni, e sistemi a regole, la logica computazionale, l'apprendimento automatico (anche probabilistico) e l'analisi dei dati. Nel corso della sua attività scientifica ha affrontato aspetti teorici, realizzativi e applicativi, come documentano le sue pubblicazioni. I suoi dati bibliometrici in termini di citazioni ed indici, da Scopus, sono: H-index 20, 201 pubblicazioni, 1662 citazioni.

E' stata Principal Investigator (coordinatore di unità) in progetti di ricerca, nazionali e internazionali, tra cui: EU Global Computing Action Project IST-2001-32530 (SOCS), progetto europeo del V programma Quadro, su logica computazionale per specifica e verifica di società di agenti software (2001-2005); Progetto di Ricerca Finalizzata 2010 n. RF-2010-2310570 "Decision support System to improve appropriateness of repeated execution of laboratory examinations", finanziato da Ministero della Salute, capofila AUO-Ferrara, (2012-2015).

Ha partecipato, come membro di unità o Principal Investigator per i più recenti, a numerosi progetti di trasferimento tecnologico che hanno portato alla realizzazione di sistemi in uso, tra cui:

- ADES, per lo sviluppo di un sistema esperto di progettazione di impianti di segnalamento ferroviario, in uso in SASIB s.p.a., ora Aalstom;
- DNSev, un sistema esperto per validazione di esami biochimici (Dianoema – ora Noemalife S.p.a.-, progetto finanziato da MIUR), utilizzato in numerosi laboratori biochimici di ospedali italiani.

E' responsabile di un progetto di trasferimento tecnologico "Analisi automatica di Big Data per Industria 4.0", finanziato con fondi della Camera di Commercio della Provincia di Ferrara, 2018, in corso, che prevede l'applicazione di machine learning per analisi intelligente di (big) data industriali provenienti dal settore manifatturiero (due aziende partecipanti). Ulteriori informazioni: <http://docente.unife.it/evelina.lamma>.

L'Università di Ferrara interverrà nella fase di sviluppo e applicazione di soluzioni di Machine Learning (ML) per estrarre conoscenza dalla mole di dati raccolti riguardanti sia le pratiche sia gli interlocutori con l'obiettivo di creare un sistema di classificazione automatica (prodotto attraverso tecniche di ML, partendo dai dati etichettati) che al presentarsi di una nuova pratica (o di un nuovo debitore), sia in grado di restituire una misura prevista del grado di successo nell'affrontare tale pratica. Affiancherà quindi i partner progettuali nelle fasi di identificazione delle variabili (attributi) rilevanti per le varie tipologie di debitore, di pratiche e tipologia di debito e di identificazione dei sistemi di apprendimento automatico (ML) più idonei per identificare similarità e/o per produrre un classificatore automatico da applicare su nuovi casi. La consulenza riguarderà anche lo sviluppo di dimostratori e la loro convalida e verifica sperimentale.

La consulenza sarà sviluppata nell'ambito degli OR assegnati a Fire e all'Università di Messina, per un costo complessivo previsto di 200.000,00 €.

Le attività affidate all'Università di Ferrara si rendono necessarie per la realizzazione del progetto e il raggiungimento degli obiettivi, in considerazione della elevata specializzazione nel settore oggetto della consulenza e di competenze specifiche che il partenariato nel suo complesso non ricomprende.

La percentuale dell'ammontare complessivo delle spese del progetto a carico di ciascun Organismo di ricerca sono di seguito riportate:

Organismo di ricerca	Ruolo	Spese previste in €	% su totale progetto
Università di Messina	Soggetto co-proponente	903.721,60	18,11
Università di Ferrara	Consulente	200.000,00	4,01

Descrizione GTS Consulting

La GTS Consulting SRL ha elaborato l'idea progettuale e curato la progettazione dell'iniziativa oggetto della presente proposta, individuando e coinvolgendo i partner industriali e l'organismo di ricerca leader nel settore di riferimento. La GTS Consulting è una società di consulenza leader nell'ideazione, progettazione e coordinamento di progetti di ricerca e sviluppo.

Tra i più recenti progetti di R&S di cui GTS Consulting ha curato l'ideazione, la progettazione e il coordinamento:

- “GOFORIT – L'esperienza prima della partenza” PON R&I 2014-2020 – Avviso D.D. 1735 del 13/07/2017 (importo progetto 10.000.000 €, in corso di svolgimento),
- “ACQUAFIT – Efficientamento energetico e fitodepurazione per il ciclo delle acque” PON R&I 2014-2020 – Avviso D.D. 1735 del 13/07/2017 (importo progetto 10.000.000 €, in corso di valutazione),
- “SORT Sviluppo di sistemi tecnologici innovativi integrati per lo Spacchettamento, l'ORGanizzazione delle scorte e il Tracciamento dei prodotti alimentari sprecati finalizzati alla loro valorizzazione urbano”, Programma Smart Cities Nazionali - Avviso D.D. 391/Ric del 5 luglio 2012 MIUR (importo progetto 13.100.000 €, concluso),
- Progetto “ABSIDE Sottoprogetti Aquasystem - BE&Save – Siglod”, PON R&C 2007-2013 Reg. Convergen. Programma Smart Cities and Social Innovation MIUR (importo progetto 48.200.000 €, concluso),
- “PAN Lab – Progetto di potenziamento strutturale dei laboratori dell'Università di Messina per analisi degli alimenti, studio della loro incidenza sulla salute umana e consulenza tecnologica, giuridica ed economica alle aziende agroalimentari”, PON “R&C 2007-2013” Regioni Convergenza Avv. n. 254/Ric MIUR (importo progetto 22.500.000 €, concluso),
- “STEM STELO - Sistemi e TECnologie per la realizzazione di Macchine per lo Sviluppo dei Trasporti Eccezionali e della LOGistica di progetto”, PON “R&C 2007-2013” Regioni Convergenza Avviso D.D. 01/Ric MIUR (importo progetto 10.578.339,00 €, concluso),
- “SIBAR - Sistemi di Irrigazione Biodegradabili per l'impiego di Acque Reflue urbane per le produzioni agroalimentari” PON “R&C 2007-2013” Regioni Convergenza Avviso D.D. 01/Ric MIUR (importo progetto 5.312.983,00 €, concluso).

Collaborazioni internazionali dell'Università di Messina

Il Dipartimento di Economia dell'Università di Messina possiede le competenze e le esperienze specifiche e l'attinenza rispetto alle tecnologie al cui sviluppo è finalizzato il progetto presentato nell'ambito del bando MISE PON “Imprese e Competitività” 2014 – 2020 - Fondo crescita Fabbrica Intelligente, Agrifood e Procedura a sportello DD 20 novembre 2018, come attestato dalle attività svolte nell'ambito della ricerca industriale ovvero dello sviluppo sperimentale dei progetti di ricerca finanziati all'Università di Messina e di cui il Dipartimento di Economia è stato partner.

In particolare il FIRE. PHA (Fire Centre for Research on Payment Habits Analysis) per lo svolgimento delle sue attività si avvale della collaborazione di altri docenti strutturati presso l'Università di Messina nonché presso altre istituzioni accademiche e centri di ricerca italiani e stranieri. Per quanto attiene al settore dell'economia, le collaborazioni internazionali coinvolgono le seguenti istituzioni: Imperial College (London, UK), King's College (London, UK), Queen Mary University (London, UK), TempleUniversity (Philadelphia, US) e University of Pennsylvania (Philadelphia, US). Per quanto attiene al settore del diritto, sono state attivate le collaborazioni internazionali con le seguenti istituzioni: University of Oxford (Oxford, UK), University of Cambridge (Cambridge, UK), University of London (Brunel College), Universidad CEU San Pablo (Madrid, Spagna) e Universität Hamburg (Amburgo, Germania).

2 QUALITÀ DELLA PROPOSTA PROGETTUALE

2.1. FATTIBILITÀ TECNICA DEL PROGETTO

Adeguatezza delle risorse strumentali e organizzative

Risorse strumentali Le risorse strumentali necessarie sono quelle informatiche ordinarie, di cui si prevede il normale ammodernamento.

Risorse organizzative – procedure L'attività sarà organizzata secondo un processo integrato che si articola su quattro fasi, concepite in modo da garantire la flessibilità necessaria nei rapporti tra i proponenti e, al contempo, la qualità conseguente da un rigoroso impiego degli strumenti di organizzazione del lavoro: (1) pianificazione del lavoro, (2) azioni comuni, (3) monitoraggio dei risultati, (4) controllo ed aggiornamento delle scelte programmatiche.

Il responsabile scientifico del progetto pianificherà il lavoro mediante tre strumenti: il cronoprogramma esecutivo (aggiornato sulla base del risultato delle fasi di monitoraggio e controllo e inizialmente redatto sulla base delle azioni previste), il quadro sinottico delle attività e la matrice ruoli/responsabilità interna ai diversi soggetti proponenti (anch'essa soggetta ad aggiornamenti) e il piano della qualità.

Questo approccio metodologico garantisce: (1) un'adeguata flessibilità di pianificazione e gestione del progetto, che consente di apportare le necessarie rimodulazioni della pianificazione in relazione a esigenze emergenti in corso d'opera e/o alle esigenze degli attori coinvolti e di garantire un'ottimale circolazione delle informazioni; (2) un adeguato supporto con procedure per il controllo di qualità del servizio.

L'assunzione di un "Sistema di qualità del progetto", sintetizzato in un piano della qualità, rappresenta lo strumento operativo di riferimento per garantire una maggiore efficienza nella governance del progetto, assicurare la verifica del conseguimento dell'obiettivo generale del progetto e verificare la rispondenza delle attività prestate e degli output realizzati rispetto alle aspettative dei soggetti proponenti e alle previsioni di capitolato.

Il piano della qualità conterrà dati di capitolato, quali: (1) la descrizione dei principali obiettivi del progetto, degli input, degli output e dei risultati attesi e dell'organizzazione di progetto; (2) la descrizione della metodologia di progetto con il dettaglio delle attività previste, la loro organizzazione gerarchica e/o temporale, e la matrice ruolo/responsabilità. Esso conterrà inoltre: (3) l'elenco della normativa, delle procedure e degli standard applicabili al progetto; (4) le procedure interne di controllo (PIC) della qualità da prevedersi nell'ambito del progetto; (5) le procedure per l'elaborazione, la trasmissione e l'approvazione formale dei documenti di progetto; (6) le procedure per garantire la riservatezza dei dati, dei documenti e delle informazioni acquisite nel corso dell'esecuzione delle attività.

Le procedure sopra citate saranno elaborate in accordo tra i proponenti in modo da rendere il più possibile oggettivo, sistematico e verificabile lo svolgimento delle attività e definiranno, in particolare: (1) responsabilità e relazioni all'interno del gruppo di lavoro che gestisce, sviluppa e verifica una data attività; (2) modalità con cui le differenti attività devono essere svolte; (3) documentazione che deve essere utilizzata ed emessa e le relative procedure di registrazione.

La validità delle soluzioni organizzative e il livello di rispondenza del procedimento alle specifiche definite saranno oggetto di verifica mediante l'adozione di indicatori di performance che saranno utilizzati per: (1) evidenziare il grado di efficienza delle prestazioni e il livello complessivo di efficacia del procedimento con particolare riferimento ai tempi di risposta, al rispetto delle scadenze e alla qualità dei risultati ottenuti; (2) evidenziare se i servizi di collaborazione forniti soddisfano i criteri di completezza e competenza richiesti; (3) caratterizzare il grado di soddisfazione del rapporto di collaborazione e i fattori di miglioramento, attraverso indicatori di gradimento.

All'analisi degli indicatori, ove si rilevino scostamenti dai valori attesi, farà seguito l'individuazione delle cause, la valutazione di soluzioni alternative e la formulazione di proposte di miglioramento.

Risorse organizzative – esperienza e competenza del responsabile scientifico Il responsabile del progetto è Maurizio Brocato, Ing., PhD, HDR, ha una consolidata esperienza nella gestione di progetti di ricerca e sviluppo, universitari e industriali. Professore universitario in Francia, dove insegna nella “École nationale supérieure d'architecture Paris-Malaquais” e nella “École nationale des ponts et chaussées”, è titolare dal 2002 di un'abilitazione a dirigere le ricerche (titolo indispensabile nel sistema della ricerca universitaria francese per svolgere questo ruolo), ed è direttore del laboratorio di ricerca “GSA – Géométrie Structure Architecture”, tra i più importanti in Francia per lo studio delle interazioni fra architettura e scienze esatte.

Le sue attività di ricerca riguardano o hanno riguardato temi quali la morfologia e l'ottimizzazione strutturale e l'innovazione tecnologica a esse associata. Questi sono stati trattati dal punto di vista teorico, senza perdere di vista le applicazioni, giungendo alla realizzazione di prototipi basati su nuove metodologie di analisi computazionali e nuove tecnologie costruttive.

Nel corso della carriera, M. Brocato ha partecipato a 22 progetti industriali di R&S, prevalentemente entro i Programmi Quadro Europei, scritto o diretto oltre 50 rapporti di R&S.

Nel contesto della presente proposta, egli potrà fare valere la propria esperienza di gestione dei progetti di ricerca, dando la giusta priorità al rispetto delle scelte programmatiche di sopra esposte e degli strumenti di programmazione corrispondenti (Gantt, matrice dei ruoli e delle responsabilità), senza perdere di vista la necessaria flessibilità delle azioni di ricerca rispetto a eventuali sopravvenute esigenze.

Dal punto di vista scientifico, il responsabile potrà far valere la propria conoscenza dei metodi di modellazione di sistemi complessi e quella dei metodi numerici computazionali, in particolare in sistemi di ricerca multi-obiettivo.

La lista che segue presenta i progetti di ricerca finanziati che il responsabile scientifico ha diretto negli ultimi dieci anni.

- 2018-20 / “GOFORIT L'Esperienza prima della partenza”, MIUR Avviso 1735.
- 2017-19 / “L'usage innovant de la pierre dans l'architecture contemporaine en Palestine (2)”, Campus France, Programme Hubert Curien Al Maqdisi 2017.
- 2016-19 / “SORT – Development of innovative and integrated technological systems of unpackaging, stock managing, and tracking, for the valorisation of food waste”, MIUR PON 2007-13 Avviso 391/Ric 05/07/2012 “Smart cities and social innovation”.
- 2016-17 / “AVACAT – Vulnerability analysis of capetian architecture facing storms”, CNRS and University Paris-Est within the PEPS call 2016.
- 2014-15 / “L'usage innovant de la pierre dans l'architecture contemporaine en Palestine (1)”, Campus France, Programme Hubert Curien Al Maqdisi 2014.
- 2013-15 / “SAVE – Technologies and operational models for the sustainable management of the food chain through the energy recovery of biological waste, the reduction of food waste in the distribution system and consumers, and the treatment and the recovery of the edible fraction of municipal solid waste”, MIUR PON 2007-13 Avviso 84/Ric “Smart cities and social innovation”.
- 2013-14 / “Team Paris”, University Paris-Est. Project “Liv-lib” built in Versailles Cité du Soleil during the Solar Decathlon Europe 2014 contest.
- 2011-14 / “SIBAR, Biodegradable Irrigation Systems Usable with UrbanWasteWater for Agroindustrial Productions”, MIUR PON 2007-13 – “Interventi di sostegno della ricerca industriale”.
- 2011-14 / “STEM-STELO, Systems and Technologies for the Production of Machines for the Development of Heavy Transportations and Project Logistics”, MIUR PON 2007-13 – “Interventi di sostegno della ricerca industriale”.

- 2011-14 / “NTFSIIE, New Photovoltaic Technologies for Building Integrated Smart Systems”, MIUR PON 2007-13 – “Interventi di sostegno della ricerca industriale”.
- 2011-14 / PhD CIFFRE contract, Association Nationale de la Recherche et de la Technologie (France), company Société Nouvelle Bâtiment Regional (France).
- 2010-13 / [Re][For][Me] French Ministry of Culture and Communication, call “Ignis mutat res 2010”.
- 2010-12 / “astonyshine” Team, ENSA Paris-Malaquais, École des Ponts, Polytechnic School of Bari and University of Ferrara. Projet “astonyshine”, built in Madrid Villa Solar during the Solar Decathlon Europe 2012 contest.

Pertinenza dei costi e la congruità delle attività progettuali I costi esposti per la realizzazione del progetto sono ritenuti congrui e pertinenti. Per le ore del personale si è tenuto conto dell’impegno previsto in relazione alle attività previste; per le attrezzature si è ritenuto di dover potenziare la dotazione ITC per poter agevolmente sviluppare le attività di progetto. Le Consulenze sono ritenute necessarie in quanto apportanti competenze altamente specialistiche che il partenariato nel suo complesso non possiede. I Materiali infine sono necessari allo sviluppo delle attività, sia nella Fase di Ricerca che nella Fase di sviluppo e test.

Nella pianificazione temporale delle attività, infine, si è tenuto sia delle interconnessioni tra i diversi OR che degli eventuali vincoli. Considerando poi le attività di cui ogni OR è composto, la programmazione dell’intera attività progettuale consente il raggiungimento degli obiettivi nei tempi prefissati.

2.2. RISULTATI ATTESI

Il risultato finale del progetto, ovvero la realizzazione di un sistema di supporto alle decisioni che ottimizzi l’equità sociale e l’economia delle azioni di recupero credito – denominato MIRC.0 – è assolutamente originale e innovativo. Per valutarne l’utilità, è sufficiente ripercorrere le note sul contesto presentate nella II^a Parte di questo documento, capitoli 4 e 5, che forniscono – in sintesi – questo scenario:

- Il potere d’acquisto delle famiglie italiane è minacciato, da anni, da bassi livelli di reddito, precarietà, disoccupazione e elevate passività, soprattutto se si fa riferimento al terzo meno ricco della popolazione;
- Il credito al consumo è, in questo contesto, una risorsa fondamentale per il Paese, che va tutelata, tenendo conto dei fattori individuali e sociali che pesano sui cittadini / utenti / consumatori;
- Plausibili scenari futuri per il nostro Paese, già in atto in altre economie occidentali, vedono l’indebitamento individuale come un sistema economico strutturale, che dovrà essere gestito per salvaguardare i cittadini e la collettività.

Gli enti di tutela del credito hanno, oggi, un approccio human driven che, se permette di valorizzare la capacità degli operatori umani di comprendere risvolti complessi e talora nascosti dei problemi per proporre soluzioni ragionevoli, pecca nella possibilità di poter conoscere e gestire informazioni complesse riguardo l’oggetto, nella rapidità di elaborazione e nella possibilità di operare controlli adatti al contesto. Questa metodologia si trova probabilmente al limite delle sue potenzialità, in termini di tempi e costi per la gestione delle pratiche, per cui è difficile immaginare uno sviluppo sociale come quello delineato nelle premesse che sia accompagnato da un sistema di tutela del credito non sostenuto da apposite ICT.

Il progetto punta conseguentemente a una traslazione – graduale – del sistema verso un approccio data driven, iniziando il processo con la creazione di un sistema che sia di supporto alle decisioni degli operatori umani. Un simile sistema non è attualmente disponibile sul mercato e i proponenti non sono al corrente di sviluppi analoghi presso altre imprese del settore.

Particolarmente importante è, nella creazione del sistema di supporto alle decisioni, il ruolo che rivestiranno in questo progetto le tecnologie informatiche di Machine Learning e di Natural Language Processing. La loro applicazione al contesto di riferimento ha un grande contenuto innovativo.

Come indicato al §5.1 della II^a Parte di questo documento, l'obiettivo finale del progetto prevede che il sistema MIRC.0 sia in grado di fornire tre prestazioni di impatto crescente sullo shift dall'approccio attuale human driven al futuribile data driven:

1. Gestione del rischio, che permette di individuare le pratiche a più alta probabilità di soddisfazione del credito per ridurre la dispersione delle risorse umane;
2. DSS propriamente detto, che supporti gli operatori ai vari livelli di gestione delle pratiche;
3. Interfaccia conversazionale, capace di gestire un primo contatto con gli interlocutori esterni.

Per quanto riguarda, più in dettaglio, i risultati attesi delle singole attività, questi sono enumerati alla fine delle stesse e leggibili sul cronoprogramma del capitolo 10 della II^a. Se ne riporta qui di seguito la lista per facilità di lettura.

- OR1.1 Identificazione dei dati, etichettatura dei dati e tipologie di analisi
 1. Prima identificazione dei dati (strutturati) disponibili e loro etichettatura: $T0 + 6$ mesi
 2. Seconda identificazione dei dati (strutturati) disponibili e identificazione di più etichettature utili al ML: $T0 + 12$ mesi
 3. Identificazione finale dei dati (strutturati, e dati estratti dalle conversazioni) utili al ML: $T0 + 24$ mesi
- OR1.2 Algoritmi di ML non supervisionato
 1. Prima identificazione degli algoritmi/strumenti più utili: $T0 + 9$ mesi
 2. Prima applicazione degli algoritmi/strumenti ai dati: $T0 + 18$ mesi
 3. Retroazione e seconda applicazione degli algoritmi/strumenti ai dati: $T0 + 30$ mesi
- OR1.3 Algoritmi di ML supervisionato
 1. Prima identificazione degli algoritmi/strumenti più utili: $T0 + 9$ mesi
 2. Prima applicazione degli algoritmi/strumenti ai dati etichettati: $T0 + 18$ mesi
 3. Retroazione e seconda applicazione degli algoritmi/strumenti ai dati: $T0 + 34$ mesi
- OR2.1 Creazione data base integrato
 1. Versione 1 del data base integrato: $T0 + 6$ mesi;
 2. Versione 2: $T0 + 18$ mesi;
 3. Versione finale: $T0 + 30$ mesi.
- OR2.2 Principali determinanti del default
 1. Prima serie di determinanti del default: $T0 + 12$ mesi
 2. Serie finale: $T0 + 24$ mesi.
- OR2.3 Principali Determinanti della LGD
 1. Prima serie di determinanti della LGD: $T0 + 18$ mesi;
 2. Seconda serie: $T0 + 27$ mesi;
 3. Serie finale: $T0 + 36$ mesi.
- OR3.1 Definizione degli esempi per convalida e verifica sperimentale dei dimostratori
 1. Identificazione “esempi” per convalida e verifica sperimentale dei dimostratori: $T0 + 12$ mesi

- OR3.2 Sviluppo e applicazione dei dimostratori
 1. Definizione GUI dei dimostratori: $T0 + 12$ mesi ;
 2. Integrazione algoritmi ML nei dimostratori e prima validazione: $T0 + 24$ mesi
 3. Retroazione, e seconda validazione sperimentale utilizzando anche informazioni estratte dalle conversazioni relative agli esempi da OR4 e OR5: $T0 + 36$ mesi
- OR4.1 Creazione Corpus file audio e documenti taggati
 1. Versione 1 del corpus files Audio e del corpus documenti: $T0 + 6$ mesi;
 2. Versione 2: $T0 + 12$ mesi;
 3. Versione finale: $T0 + 30$ mesi.
- OR4.2 Riconoscimento file audio (ASR)
 1. Primo Corpus file trascritti (circa 10%): $T0 + 6$ mesi
 2. Secondo Corpus file trascritti (circa 50%): $T0 + 12$ mesi
 3. Corpus di test completo (100%): $T0 + 24$ mesi.
- OR4.3 Analisi linguistica profonda dei testi
 1. Versione 0: $T0 + 12$ mesi;
 2. Versione 1: $T0 + 24$ mesi;
 3. Versione finale: $T0 + 34$ mesi.
- OR5.1 Analisi del contesto
 1. Definizione e organizzazione dei gruppi di interesse (stakeholder): $T0 + 6$ mesi;
 2. Output dell'analisi: $T0 + 18$ mesi.
- OR5.2 Sviluppo del sistema
 1. Versione 0: $T0 + 24$ mesi;
 2. Versione finale: $T0 + 30$ mesi.
- OR5.3 Test sul sistema sviluppato e feedback
 1. Test predisposti: $T0 + 30$ mesi;
 2. Ultimo feedback: $T0 + 34$ mesi.
- OR5.4 Integrazione con gli altri moduli
 1. Sistema integrato in MIRC.0: $T0 + 36$ mesi.
- OR6.1 Progettazione interfaccia grafica
 1. Design della GUI: $T0 + 6$ mesi;
 2. Primi test con utenti: $T0 + 18$ mesi;
 3. Versione finale: $T0 + 36$ mesi.
- OR6.2 Sviluppo ambiente dimostrativo e di test continuo dell'analisi linguistica
 1. Versione 0: $T0 + 30$ mesi;

2. Versione finale: $T0 + 36$ mesi.
- OR6.3 Test con diverse categorie di utenti
 1. Versione 0: $T0 + 30$ mesi;
 2. Versione finale: $T0 + 36$ mesi.
 - OR7.1 Definizione delle basi dati da utilizzare
 1. Elenco basi oggettive e accesso a queste basi: $T0 + 6$ mesi;
 2. Elenco basi statistiche o altro e accesso a queste basi: $T0 + 12$ mesi.
 - OR7.2 Pulizia dei dati
 1. Organizzazione delle procedure per la base dati Fire $T0 + 12$ mesi;
 2. Sistema di pulizia per basi generiche: $T0 + 18$ mesi.
 - OR7.3 Sicurezza dei dati
 1. Scelta del sistema $T0 + 21$ mesi;
 2. Sistema funzionante: $T0 + 30$ mesi.
 - OR7.4 Oscuramento dei dati
 1. Scelta del sistema $T0 + 27$ mesi;
 2. Sistema funzionante: $T0 + 36$ mesi.
 - OR8.1 Metodologie di base
 1. Organizzazione dell'OR: $T0 + 4$ mesi;
 - OR8.2 Analisi della conoscenza dei phone collector
 1. Output 1: $T0 + 10$ mesi;
 2. Output 2: $T0 + 16$ mesi;
 3. Output 3: $T0 + 22$ mesi;
 4. Output finale: $T0 + 28$ mesi.
 - OR8.3 Analisi della conoscenza degli home collector
 1. Output 1: $T0 + 11$ mesi;
 2. Output 2: $T0 + 17$ mesi;
 3. Output 3: $T0 + 23$ mesi;
 4. Output finale: $T0 + 29$ mesi.
 - OR8.4 Analisi della conoscenza dei legal collector
 1. Output 1: $T0 + 12$ mesi;
 2. Output 2: $T0 + 18$ mesi;
 3. Output 3: $T0 + 24$ mesi;
 4. Output finale: $T0 + 30$ mesi.
 - OR8.5 Analisi della conoscenza dei production manager

1. Output 1: $T0 + 13$ mesi;
 2. Output 2: $T0 + 19$ mesi;
 3. Output 3: $T0 + 25$ mesi;
 4. Output finale: $T0 + 31$ mesi.
- OR9.1 Metodologie di base
 1. Organizzazione dell'OR: $T0 + 11$ mesi;
 - OR9.2 Test semi-automatici
 1. Definizione dei test: $T0 + 24$ mesi;
 2. Test pronti per l'esecuzione: $T0 + 28$ mesi;
 3. Fine test: $T0 + 34$ mesi;
 - OR9.3 Test con simulazione
 1. Definizione dei test: $T0 + 24$ mesi;
 2. Test pronti per l'esecuzione: $T0 + 28$ mesi;
 3. Fine test: $T0 + 34$ mesi;
 - OR9.4 Analisi diretta
 1. Istruzione operatori: $T0 + 24$ mesi;
 2. Test pronti per l'esecuzione: $T0 + 28$ mesi;
 3. Fine test: $T0 + 34$ mesi;
 - OR9.5 Interazione con agenti esterni
 1. Definizione dei contatti: $T0 + 26$ mesi;
 2. Test pronti per l'esecuzione: $T0 + 30$ mesi;
 3. Fine test: $T0 + 36$ mesi;

2.3. GRADO DI INNOVAZIONE

La tipologia di innovazione introdotta impatta su un particolare sistema di produzione di servizi che è quello della tutela del credito. Si tratta, come indicato precedentemente, di innovazione di prodotto, in quanto il sistema di supporto delle decisioni MIRC.0 costituirà un primo esempio nel suo genere, capace di indurre un'innovazione di processo, configurando nuovi processi produttivi, sempre più basati su approcci data driven.

Le attuali procedure aziendali per il recupero dei crediti prevedono una serie di contatti presi da personale addetto con interlocutori esterni, debitori secondo una data pratica. Lo sviluppo della trattativa è interamente condotto sulla base di interazioni tra questi attori umani. Il personale deve spesso interagire con livelli gerarchici superiori per ottenere consigli o autorizzazioni. Si verificano dunque lentezze, incertezze ed eventuali errori, oltre a una ridotta capacità di accesso ai dati, sia in termini di velocità che di dimensione delle basi.

Queste limitazioni sono superabili attraverso il ricorso a ICT, per cui è possibile introdurre dei cambiamenti tecnologici radicali nel processo produttivo sopra indicato, migliorandolo notevolmente. D'altro canto è unica e indispensabile nel contesto la capacità delle risorse umane di gestire situazioni complesse e di operare scelte dinanzi a dati parziali, informazioni incomplete, possibilità di errore e difficoltà di applicazione di metodi predittivi certi. Pertanto si vuole realizzare un sistema informatico che potenzi le capacità degli operatori umani (senza rimpiazzarli) e ne migliori al tempo stesso le capacità produttive per l'azienda e le capacità di interazione con gli interlocutori esterni per un più equo svolgimento delle trattative.

Un processo di recupero crediti nel quale si possa:

MIRC.0

- Tenere conto in modo quanto più puntuale possibile delle reali esigenze dell'interlocutore;
- Evitare perdite di tempo nel corso delle trattative;
- Evitare trattative inutili;

costituisce un grande vantaggio competitivo per l'azienda che lo possiede e un'indubbio miglioramento del processo e degli effetti che questo può avere nei confronti degli interlocutori.

3 IMPATTO DEL PROGETTO

3.1. INTERESSE INDUSTRIALE

Il contesto industriale di riferimento è quello della tutela del credito. Come più volte indicato in questo documento, si tratta di un'attività che ha importanti ripercussioni sul sistema economico e sociale del Paese: tutelare il credito significa al tempo stesso garantire flussi di cassa verso le aziende (soprattutto quelle orientate verso il mercato interno) e le amministrazioni e fornire potere di acquisto ai cittadini. Il credito è infatti, per sua natura, la valorizzazione nel presente di risorse economiche future, operazione che assume un alto valore aggiunto tanto in un contesto – come l'attuale italiano – di scarsa crescita economica, quanto in uno – osservabile soprattutto in altri Paesi OCSE ed eventualmente prevedibile nel nostro – di forti dinamiche salariali.

Nelle parti precedenti di questo documento sono già state descritte alcune caratteristiche di questo contesto; si riportano qui di seguito alcuni passaggi utili per capirne l'interesse industriale. Essi vanno letti avendo in mente che si tratta di un interesse che travalica quello dell'azione imprenditoriale, impattando su dinamiche socio-economiche che dovrebbero sollevare l'attenzione del legislatore.

E' convinzione della compagine proponente che il progetto MIRC.0 sia innanzi tutto un'ipotesi di ristrutturazione del rapporto debitorio tra cittadini e enti erogatori di beni e servizi che può fare transitare la nostra società verso maggiori livelli di equità, soprattutto grazie all'applicazione di ICT.

In Italia, il mercato dei crediti da recuperare ha fatto registrare una crescita più che significativa, sebbene esistano segmenti ancora pressoché vergini (si pensi ai crediti commerciali): la concentrazione intorno agli NPL rischia di far perdere focus sul più tradizionale credito al consumo e sulle utenze. E' certo che in generale tutti i comparti offrono oggi crediti da recuperare molto più complessi, pur a fronte di commissioni sempre meno interessanti per l'impresa di recupero del credito. C'è dunque una doppia sfida di fronte alla quale si trova l'intero comparto: far percepire il recupero dei crediti per ciò che è in realtà, ossia un regolatore sociale, una fonte di risparmio di risorse per le aziende che devono accettare di sostenere un costo superiore; la seconda sfida è intervenire sui propri mezzi di produzione, costruendo un nuovo paradigma operativo.

Il processo di recupero crediti può realizzare oggi un continuum operativo grazie alla completa automazione delle attività di import, scoring e gestione ad eventi: l'anima di questo percorso è costituita dal cluster engine, un workflow realizzato dalla funzione di Information Technology che – oggi – muove le posizioni da gestire tra gli operatori a seconda degli eventi che si manifestano, applicando regole personalizzate per ciascun cliente (ente creditore), per le diverse anzianità dei crediti.

La personalizzazione del rapporto con gli interlocutori (individui debitori) è invece più lontana dall'essere, a causa della molto più grande diversità dei soggetti e delle condizioni di difficoltà finanziaria che possono aver giustificato l'interruzione del rapporto creditizio inizialmente contratto. Si tratta di una debolezza del sistema che può essere rimossa, trasformandosi in opportunità.

Oggi, l'IA consente infatti di andare oltre una gestione puramente "numerica" dei dati, aggiungendo la capacità di comprendere e gestire le informazioni per ricavarne il massimo valore di business. La sua applicazione al contesto dell'interlocutore in una pratica di recupero può dunque condurre a nuove strategie, basate sulle grandi moli di dati che caratterizzano questo contesto.

Nel mondo moderno, è di vitale importanza per le organizzazioni di qualsiasi dimensione disporre della capacità di creare, catturare, archiviare, individuare e condividere la conoscenza, ma anche disporre di adeguate piattaforme che consentano di usufruire delle informazioni necessarie al momento giusto. La tecnologia semantica, con la sua capacità analoga a quella della mente umana di capire il significato delle parole in base al contesto, può supportare individui ed organizzazioni nel cruciale processo di gestione

delle informazioni disponibili, e sempre più spesso consente alle imprese di rispondere rapidamente ai cambiamenti del business, ottenere informazioni approfondite su clienti, industrie e mercati, e diventare soggetti più proattivi.

Per rispondere a questa sfida, come indicato nelle parti precedenti di questo documento, il sistema MIRC.0 che si intende realizzare dovrà fornire prestazioni a tre livelli possibili nel contesto operativo del recupero crediti:

Prestazione 1: gestione del rischio Attualmente, presso Fire come in altre imprese del settore, i modelli di ottimizzazione della gestione delle pratiche di recupero crediti e di definizione del profilo di rischio del debitore utilizzano metodi classici, basati su strumenti di regressione lineare o di associazione statistica. L'idea di questo progetto è di sfruttare le recenti tecniche di analisi e di classificazione utilizzando strumenti di machine learning (ML), per cui l'aspettativa è di migliorare notevolmente la capacità predittiva del rischio e di rendimento atteso da ogni pratica e, al contempo, migliorare l'identificazione della modalità di recupero ottimale anche per l'interlocutore. Questo miglioramento può consentire di individuare le pratiche a più alta probabilità di soddisfazione del credito e concentrare l'utilizzo di risorse umane su queste ultime, permettendo all'impresa di fare leva sulla esistente forza lavoro per incrementare l'attività di gestione. Una migliore capacità predittiva del rischio può anche consentire di spingere al di là dei livelli oggi accettabili la trattativa con interlocutori ritenuti affidabili (per esempio, individui sottoposti a momentanea pressione economica comprovata dall'incrocio di dati).

Prestazione 2: sistema di supporto alle decisioni All'interno dell'iter di gestione di ogni pratica di credito, le fasi iniziali di phone servicing sono quelle a maggior intensità di forza lavoro; esse pertanto incidono maggiormente sui margini operativi e profitti dell'impresa e sui rapporti con gli interlocutori. L'operatore telefonico deve spesso interagire con il livello gerarchico superiore e questo comporta tempi che potrebbero essere ridotti grazie all'interazione con un sistema esperto. Inoltre questa interazione tra operatori umani genera, nel contesto attuale, rischi di inefficienza, perdita di informazione, riduzione delle opzioni, che possono essere evitati col ricorso a strumenti informatici. Una seconda idea del progetto è dunque quella di realizzare un sistema esperto interattivo. Si ritiene che in un primo tempo il sistema possa limitarsi ad ascoltare le conversazioni e agire solo attraverso gli operatori umani, mediante un'interfaccia appositamente progettata, lasciando agli stessi la possibilità di decidere sul proseguimento del rapporto con l'interlocutore, per passare in un secondo tempo – gradualmente e per i passaggi più semplici – a un'interazione diretta con gli interlocutori mediante un'interfaccia robotica.

Prestazione 3: interfaccia conversazionale La terza idea del progetto propone di utilizzare le recenti trasformazioni tecnologiche nel settore della robotica della comunicazione per cercare di automatizzare e robotizzare quanto più possibile la fase di phone servicing.

Innovazioni perseguite Attraverso l'introduzione di un prodotto ICT innovativo, il progetto giunge a definire un nuovo modello di business per la gestione del credito, cioè un'innovazione di processo, che sposta il focus da un approccio human driven a uno sempre più data driven. Per l'interlocutore questo spostamento può condurre all'ampliamento delle garanzie che alcune reali condizioni di sofferenza possano essere recepite, visto che il sistema può avere accesso a una mole molto più importante di informazioni dell'operatore umano e l'interlocutore e la pratica che lo riguarda possono essere profilati in modo molto più completo e complesso. Per l'impresa, i cui profitti si definiscono in termini di margini sul ricavato, un focus data driven permette di spalmare la forza lavoro esistente su un numero maggiore di pratiche e di focalizzarla su quelle a più alta probabilità di successo, garantendo un impatto positivo sui profitti.

Relativamente all'innovazione di prodotto – che è primaria per il successo dell'iniziativa, il progetto punta alla creazione di un nuovo sistema di supporto alle decisioni (che opera in modo totalmente diverso da quanto normalmente fatto dai software di credit scoring), avente tre funzioni principali: a) tenere conto di un profilo complesso dell'interlocutore e della pratica che lo riguarda allo scopo di venire incontro alle sue reali esigenze; b) suggerire percorsi ottimali, per ciascuna pratica di credito, e valutare la probabilità di soddisfazione del credito stesso, corredando questa informazione con la durata prevista dell'intera gestione della pratica in funzione dei percorsi alternativi; c) pervenire alla determinazione di un costo di ciascuna pratica (o di un lotto di pratiche, tenendo presente le correlazioni che potenzialmente ne

riducono il rischio). L'anticipazione del costo consente sia di rivelare all'impresa che tipo di margini applicare per gestire la singola oppure il lotto di pratiche, sia di rivelare un fair value, utile nel caso la Fire decidesse di assumere in proprio il rischio di non evasione della stessa.

3.2. POTENZIALITÀ DI SVILUPPO

Un recente rapporto sulle frontiere dell'IA in ambito industriale ⁽¹⁾ Mostra come le applicazioni di IA, soprattutto se basate su ML, non soltanto forniscono un crescente contributo all'economia industriale, ma sono fondamentali nella maggior parte dei casi nei quali intervengono per la soluzione di un problema. Si tratta dunque di tecnologie di punta estremamente adatte agli sviluppi industriali.

Tra le applicazioni di maggior successo si trovano quelle che puntano alla produzione di servizi ad hoc (customizzati) per il consumatore. Si può notare che i servizi bancari e le assicurazioni (che tra i servizi studiati sono quelli che più si avvicinano alla tipologia di interesse per questo progetto) hanno fatto contare un alto numero di casi di impiego di tecniche di IA e in particolare di ML.

Si ritiene pertanto che il progetto rappresenti un'opportunità di ingresso delle suddette tecnologie ICT nel mondo della tutela del credito che non potrà che avere effetti positivi nei prossimi anni. Si tratta infatti, per i motivi che sono stati più volte descritti in questo progetto, di un contesto di servizi fortemente guidato da un approccio human driven, nel quale – pur nell'ambito di un approccio analitico avanzato per quanto riguarda l'analisi dei rischi – si impiegano tecniche e algoritmi più tradizionali nell'approccio con gli interlocutori le cui performance sono oggi decisamente superabili.

Si aggiungano a questi aspetti tecnologici, altri che riguardano maggiormente l'attenzione alla questione sociale collegata alla tutela del credito: la salvaguardia del potere d'acquisto e del futuro delle famiglie. La possibilità di inserire nel sistema del recupero crediti azioni customizzate, appositamente concepite caso per caso, non può che migliorarne la flessibilità e avere impatti positivi sui cittadini in difficoltà economica.

Allo stato attuale, non è possibile immaginare una risoluzione ad hoc delle trattative, ma è necessario attenersi a schemi negoziati prestabiliti. Questa necessità deriva sostanzialmente dall'impossibilità di accedere a dati ed effettuare simulazioni caso per caso con sufficiente margine di precisione da poter sostenere le scelte necessarie a una risoluzione personalizzata delle trattative. Questa possibilità è invece offerta da un sistema di supporto delle decisioni opportunamente progettato.

L'idea di MIRC.0 ha dunque grandi potenzialità di sviluppo e potrebbe condurre a una radicale trasformazione del settore.

¹McKinsey Global Institute “Notes from the AI frontier. Insights from hundred of use cases” Apr. 2018.

4 RISORSE FINANZIARIE PER LA REALIZZAZIONE DEL PROGETTO

Per il soggetto proponente Fire i costi al netto di personale interno già in forza all'azienda e delle spese generali ammontano in totale a 458.001,53 €. A questi vanno sommate le spese ammissibili non ricomprese nei costi ammissibili, specificatamente la differenza tra la spesa relativa alle attrezzature e il costo ammissibile calcolato in relazione al periodo di ammortamento ed all'effettivo impegno nelle attività di progetto. Spese per Attrezzature 62.663,00 € – costo ammissibile 20.001,53 € = 42.661,47 €. Il Totale dei Fabbisogni risulta così pari a 500.663,00 €. A tale importo viene poi sommata l'IVA per 110.145,86 € (calcolata al 22%) per un totale complessivo di 610.808,86 €.

I Fabbisogni sono stati poi suddivisi per anno in coerenza con la suddivisione percentuale delle spese previste in domanda (30% primo anno – 30% secondo anno – 40% terzo anno). La quota di 42.661,47 € è stata attribuita in toto al primo anno, in quanto le attrezzature verranno acquistate entro il primo mese di avvio del progetto.

Tra le Fonti al primo anno si inserisce il finanziamento agevolato pari a 566.699,38 €, considerato che la Fire ha intenzione di richiederlo in anticipazione. Dal secondo anno si inseriscono le quote di contributo che saranno richieste calcolate in relazione ai costi che si prevede di sostenere nel periodo.

Si evidenzia come il finanziamento agevolato e le seguenti quote di contributo riescano, opportunamente modulate nel tempo, a fornire le fonti necessarie a coprire l'intero fabbisogno.

PIANO FINANZIARIO PER LA REALIZZAZIONE DEL PROGRAMMA				
FABBISOGNO	ANNO 1	ANNO 2	ANNO 3	ANNO 4
Spese previste ammissibili (1)	180.061,93	137.400,46	183.200,61	0,00
IVA	39.613,62	30.228,10	40.304,13	
Totale	219.675,56	167.628,56	223.504,75	0,00
FONTI DI COPERTURA				
Eccedenza fonti anno precedente		347.023,83	580.128,85	757.357,69
Apporto di mezzi propri/finanziamento soci				
Erogazioni finanziamento agevolato	566.699,38			
Erogazioni contributo alla spesa		400.733,58	400.733,58	534.311,44
Debiti verso fornitori (2)				
Cash-flow				
Affidamenti bancari (2)				
Totale	566.699,38	747.757,41	980.862,43	1.291.669,13

Per il soggetto proponente Expert System i costi al netto di personale interno già in forza all'azienda e delle spese generali ammontano in totale a 250.000,00 €. A questo importo viene poi sommata l'IVA per 55.000,00 € (calcolata al 22%) per un totale complessivo di 305.000,00 €. I Fabbisogni sono stati poi suddivisi per anno in coerenza con la suddivisione percentuale delle spese previste in domanda (30% primo anno – 30% secondo anno – 40% terzo anno).

Tra le Fonti al primo anno si inserisce il finanziamento agevolato pari a 250.325,00 €, considerato che la Expert System ha intenzione di richiederlo in anticipazione. Dal secondo anno si inseriscono le quote di contributo che saranno richieste calcolate in relazione ai costi che si prevede di sostenere nel periodo.

Si evidenzia come già il finanziamento agevolato riesca a fornire le fonti necessarie a coprire l'intero fabbisogno.

PIANO FINANZIARIO PER LA REALIZZAZIONE DEL PROGRAMMA				
FABBISOGNO	ANNO 1	ANNO 2	ANNO 3	ANNO 4
Spese previste ammissibili (1)	75.000,00	75.000,00	100.000,00	0,00
IVA	16.500,00	16.500,00	22.000,00	
Totale	91.500,00	91.500,00	122.000,00	0,00
FONTI DI COPERTURA				
Eccedenza fonti anno precedente		158.825,00	244.747,25	300.169,50
Apporto di mezzi propri/finanziamento soci				
Erogazioni finanziamento agevolato	250.325,00			
Erogazioni contributo alla spesa		177.422,25	177.422,25	236.563,00
Debiti verso fornitori (2)				
Cash-flow				
Affidamenti bancari (2)				
Totale	250.325,00	336.247,25	422.169,50	536.732,50