



Sviluppo portale – Valutazione dell' applicabilità di strumenti SW basati su tecnologie AI al portale SiMTE per funzioni di lettura automatica, estrazione dati, assistenza utenti. (LA1.5)

E. Cosimi, F. Donato, S. Giammartini, W. Gaggioli, A. Mariano, G. Simbolotti (ENEA)

In collaborazione con il team Sistemi Informativi/IBM coinvolto con apposito contratto per la fornitura e l'addestramento del SW-AI.

SVILUPPO PORTALE – VALUTAZIONE DELL’ APPLICABILITÀ DI STRUMENTI SW BASATI SU TECNOLOGIE AI AL PORTALE SIMTE PER FUNZIONI DI LETTURA AUTOMATICA, ESTRAZIONE DATI, ASSISTENZA UTENTI. (LA1.5)

E. Cosimi, F. Donato, S. Giammartini, W. Gaggioli, A. Mariano, G. Simbolotti (ENEA)

In collaborazione con il team Sistemi Informativi/IBM coinvolto con apposito contratto per la fornitura e l’addestramento del SW-AI.

Dicembre 2021

Report Ricerca di Sistema Elettrico

Accordo di Programma Ministero della Transizione Ecologica - ENEA

Piano Triennale di Realizzazione 2019-2021 - III annualità 2021

Obiettivo: *Sistema Elettrico*

Progetto: ‘Strumenti e modelli per scenari energetici ed elettrici adeguati all’evoluzione del sistema..., dei mercati e della regolazione’.

Linea di attività: *LA1.5 Sviluppo portale – Valutazione dell’ applicabilità di strumenti SW basati su tecnologie AI al portale SiMTE per funzioni di lettura automatica, estrazione dati, assistenza utenti.*

Responsabile del Progetto: Giorgio Simbolotti, ENEA

Indice

SOMMARIO.....	4
1 INTRODUZIONE AL PROGETTO.....	5
1.1 IMPATTO DELLA PANDEMIA COVID-19 SULLE ATTIVITÀ PREVISTE PER L'ANNO 2020.....	8
1.2 BREVI INFORMAZIONI SULLA PIATTAFORMA SIMTE.....	9
2 DESCRIZIONE DELLE ATTIVITÀ SVOLTE E RISULTATI DELLA LA 1.6.....	11
2.1 INTRODUZIONE.....	11
2.2 ATTIVITÀ SVOLTE NELLA FASE 1.....	12
2.3 RISULTATI DELLA FASE 1.....	23

Allegato 1: RELAZIONE TECNICA - Appalto per fornitura di servizi di implementazione/addestramento e diritti di accesso (license & cloud) a SW-AI (SoftWare – Artificial Intelligence) per applicazione a piattaforma SIMTE-ENEA –RTI Sistemi Informativi/IBM e VAR Group -

Allegato 2 : RAPPORTO FINALE DI ATTIVITÀ per servizi di implementazione/addestramento e diritti di accesso (license & cloud) a SW-AI (SoftWare – Artificial Intelligence) per applicazione a piattaforma SIMTE-ENEA - Sistemi Informativi/IBM

Allegato 3: Specifiche Tecniche per l'acquisto di servizi di implementazione/addestramento e diritti di accesso (license & cloud) a SW-AI (Software - Artificial Intelligence) per applicazione alla piattaforma SIMTE (Sistema informativo e di Monitoraggio delle Tecnologie Energetiche), ENEA – Agosto 2020

Sommario

Come previsto dal Piano Triennale Realizzativo (PTR) del progetto, nel corso dell'annualità 2021 sono stati svolti ed ultimati i lavori relativi alle linee di attività LA 1.5 e 1.6 che riguardano rispettivamente la prova di applicabilità e la successiva Implementazione di software Intelligenza Artificiale (SW-AI) sulla piattaforma SiMTE per svolgere le funzioni di: a) lettura automatica di documenti tecnico-economici con estrazione/gestione dei dati di interesse; b) assistenza agli utenti della piattaforma (assistente virtuale) nell'utilizzo dei servizi offerti (ricerca dati e documenti, interrogazione in linguaggio naturale, uso del modello per il calcolo di prestazioni e costi delle tecnologie energetiche ...).

L'introduzione dell'AI sulla piattaforma SiMTE è motivata dalla necessità di ridurre gli oneri (tempi e risorse) di aggiornamento della piattaforma (prestazioni e costi delle tecnologie) e di rendere più fruibili per gli utenti i servizi offerti. Essa costituisce inoltre la parte più **innovativa** del progetto

Nel lavori di introduzione dell'AI è anche confluita la LA 1.2 (Implementazione di funzioni statistiche per analisi di prestazioni e costi di tecnologie di larga diffusione commerciale) poiché le funzioni statistiche sono state introdotte con un SW aggiuntivo integrato nel sistema AI. Le tre LA 1.2, 1.5 e 1.6 danno luogo rispettivamente ai rapporti tecnici distinti 161, 163 e 071, tuttavia le LA 1.2 e 1.5 sono strettamente legate alla LA 1.6 e i rispettivi rapporti condividono parti comuni.

I ritardi riconducibili all'emergenza Covid e alcune difficoltà amministrative nell'assegnazione di contratti hanno prodotto lo slittamento di queste LA dal 2020 al 2021 (variante approvata) con inizio effettivo nell'Aprile 2021 e conseguente drastica riduzione dei tempi di lavoro da 12 a circa 6 mesi. Cionondimeno, l'introduzione dell'AI ha richiesto un numero di ore di lavoro superiore a quanto previsto a causa del laborioso processo di addestramento del SW che è stato necessario acquisire per rendere autonoma l'Agenzia nell'uso del SW stesso. Le LA 1.5 e 1.6 hanno quindi richiesto impegni superiori al preventivo iniziale, tuttavia compensati dal minor impegno richiesto per la LA 1.2 grazie alla sua integrazione nel sistema AI. Di questo si è tenuto conto in sede di rendicontazione. Nella fase iniziale del lavoro un tempo considerevole è stato inoltre richiesto per mettere in comune le conoscenze del committente in tema di tecnologie energetiche con quelle del fornitore in tema di addestramento del SW.

Risultati della LA 1.5 (prova di applicabilità): a) per svolgere le complesse funzioni richieste si è reso necessario l'uso di una versione avanzata del SW (IBM Watson), non ancora commercializzata; b) è richiesto un addestramento specifico del SW per ogni tecnologia trattata; quindi l'applicazione sistematica a tutte le 100 tecnologie trattate nella piattaforma SiMTE richiede un tempo considerevole; c) va rilevato che il processo di addestramento del SW per un certo settore tecnologico si velocizza con l'accrescimento del patrimonio cognitivo del SW (terminologia, entità, relazioni) e quindi al crescere del numero di tecnologie trattate. Il SW è stato addestrato su alcune tecnologie pilota (e.g.: fotovoltaico, solare a concentrazione) e al termine della LA 1.5 sono stati raggiunti gli obiettivi fissati dimostrando l'applicabilità dell'AI per i fini preposti. In particolare: a) acquisire documenti tecnici con estrazione dei dati di interesse (prestazioni e costi) nelle rispettive unità di misura e trasferimento alla banca dati SiMTE; b) guidare l'utente tramite assistente virtuale nella ricerca dati e interrogazione della piattaforma fornendo risposte con congruenza dell'ordine del 60% rispetto alle domande formulate; c) fornire a richiesta brevi report con testi e tabelle relativi alle domande formulate. Tutte le funzioni di cui sopra sono state provate con esiti soddisfacenti anche se al termine della LA 1.5 il sistema richiedeva naturalmente ulteriore messa a punto.

1 Introduzione al progetto

Il **Piano Triennale RdS 2019-2021** per il **Tema di Ricerca 2.1** dal titolo *“Strumenti e modelli... per scenari energetici ed elettrici adeguati all’evoluzione del sistema..., dei mercati e della regolazione...”*, assegna ad Enea i seguenti sotto-temi di ricerca:

1. *“Sviluppo di un data-base con dati tecnico-economici sulle tecnologie energetiche (prestazioni, emissioni, costi, dati di mercato, proiezioni, etc.), utilizzabili nei modelli (energetici)...”*;
2. *“Sviluppo di un portale per l’accesso a documenti e dati... con funzioni di ricerca e analisi, e modelli ... per il calcolo di prestazioni e costi delle tecnologie...”*;
3. *“Aggiornamento dei **brief** tecnologici che per ogni tecnologia contestualizzano informazioni e dati tecnico-economici attuali e attesi ...”*;
4. *“Valutazioni previsionali energetiche regionali coerenti con scenari nazionali, finalizzate a un nuovo burden-sharing delle fonti rinnovabili e del potenziale regionale di efficienza energetica ...”*.

Nei punti 1), 2) e 3) di cui sopra si fa implicito riferimento (e.g. la parola *brief*) ad una attività pregressa conclusa nel 2018 (Convenzione MISE-ENEA denominata SIMTE) nell’ambito della quale, in ottemperanza ad alcuni dispositivi di legge¹, l’Enea ha realizzato per conto del MISE un **Sistema di Monitoraggio dello stato e delle prospettive delle Tecnologie Energetiche (SIMTE)**.

Il sistema consiste in una piattaforma web (www.simte.enea.it) che offre all’utenza (prevalentemente operatori, analisti e decisori del settore energetico) documenti, informazioni, dati e strumenti di analisi sulle tecnologie energetiche commerciali e in via di sviluppo nei settori di offerta e domanda dell’energia: a) produzione e trasformazione di energia primaria; b) produzione e distribuzione di elettricità e calore; c) residenziale e servizi; d) industria; e) trasporti. Si tratta di un sistema informativo con focus sul mercato italiano e attenzione agli sviluppi internazionali, con orizzonte temporale 2018-2030 (2050) e aggiornamenti biennali (ove si renda disponibile una continuità di risorse).

Il sistema attualmente comprende:

- la piattaforma web (portale) aperta in via sperimentale all’indirizzo di cui sopra e destinata a decisori pubblici e privati, investitori e analisti del settore energia che possono accedere ai documenti e ai servizi offerti;
- circa 50 documenti tecnici (*brief*) redatti da esperti, aggiornati al Gennaio 2018, che contestualizzano dati attuali e attesi su prestazioni, emissioni, quote di mercato, normative, peculiarità nazionali, costi di investimento, di esercizio, di prodotto/servizio, etc. per altrettante tecnologie trattate;
- una banca-dati (derivata dai documenti di cui sopra e altre fonti) dotata di strumenti di ricerca, confronto e analisi dei dati;
- un modello per il calcolo di prestazioni e costi delle tecnologie in base a criteri omogenei e ad input modificabili dall’utente.

In base a quanto previsto nei punti 1-4 di cui sopra, il presente progetto si prefigge i seguenti obiettivi:

1. **Aggiornare/raddoppiare la base-dati e gli strumenti della banca-dati SIMTE**, in particolare,
 - a. aggiornare documenti e dati già presenti, il sistema di ricerca/analisi dati e il modello di calcolo prestazioni e costi delle tecnologie ad uso dell’utenza;
 - b. raddoppiare la base-dati con circa 50 nuovi documenti (*brief*) aggiuntivi su tecnologie ancora non presenti e/o tramite opportuna disaggregazione di quelle già presenti;
 - c. introdurre un database su misure di incentivazione per tecnologie rinnovabili ed efficienza energetica;

¹ DLgs 28 del 3.3.2011, c.7/art. 40, DM 6.7.2012 e DM 28.12.2012

- d. introdurre strumenti per valutazioni statistiche di prestazioni e costi di tecnologie di largo consumo (settori residenziale e trasporti), in base a dati di mercato.
2. **Conseguire un salto di qualità nella gestione della banca dati e del portale** (con sostanziale riduzione delle risorse necessarie) attraverso l'adozione di strumenti software (SW) basati su **Intelligenza Artificiale (AI)** al fine di rendere possibile:
 - a. lettura automatica di documenti tecnici con acquisizione e gestione dei dati di interesse;;
 - b. produzione automatica di report su richiesta dell'utente;
 - c. assistenza virtuale all'utenza per l'accesso ai servizi offerti dalla banca dati.
 3. **Attuare una adeguata strategia di diffusione** della piattaforma SIMTE e conseguire la piena operatività del portale verso l'utenza, i.e. operatori, decisori e investitori del settore energia, consumatori.
 4. **Produrre le valutazioni energetiche previsionali** a livello regionale coerenti con gli scenari nazionali (metodologia e applicazioni) come previsto nel sotto-tema 4.

Nella formulazione inizialmente approvata (successivamente oggetto di varianti nel 2020, vedi sez. 1.1 seguente) il Piano Triennale Realizzativo (PTR) del progetto si articola in **23 linee di attività (LA)** divise in **5 aree** e distribuite su **tre annualità**. Le 5 aree comprendono:

- Sviluppo Banca Dati, che comprende le LA 1.1, 1.2, 1.3 e 1.4;
- Sviluppo Portale, che comprende LA 1.5, 1.6 e 1.7 (con coinvolgimento di fornitore di SW-AI);
- Aggiornamento/Produzione Documenti-Brief, che comprende le LA da 1.8 a 1.16 (con coinvolgimento di partner universitari);
- Valutazioni Energetiche Regionali, che comprende le LA 1.17, 1.18, 1.19 e 1.20;
- Diffusione Risultati, che comprende le LA 1.21, 1.22 e 1.23 (con coinvolgimento di azienda specializzata in diffusione).

La prima annualità 2019 comprende 4 LA e una attività preparatoria auto-finanziata descritte di seguito:

- **LA1.1:** miglioramento del sistema di ricerca, analisi e presentazione dati; aggiornamento del modello deterministico per calcolo prestazioni/costi delle tecnologie da ambiente Excel/VB ad ambiente Scilab (open source) con inserimento di misure di incentivazione (e.g. super-ammortamento); preparazione casi-guida per analisi prestazioni/costi tecnologie;
- **LA1.8:** preparazione nuovo format per aggiornamento brief, raccolta dati e nuovi documenti;
- **LA1.17:** analisi energetiche regionali: messa a punto del metodo e applicazione a Regione Lazio;
- **LA1.21:** strategia diffusione portale (valutazioni di proposte fornite pro-bono da aziende di settore).

L'attività autofinanziata prevista per il 2019 comprende:

- **Analisi di SW basati su AI** (con la collaborazione di fornitori di SW coinvolti pro-bono) per: lettura automatica documenti, estrazione e gestione dati; redazione di brevi report su richiesta utente (meta-motori); assistenza virtuale all'utente della banca dati;

La seconda annualità 2020 comprende 12 LA e prevede il massimo impegno di risorse e investimenti con il coinvolgimento di due Università su temi di analisi e ricerca (Università di Roma la Sapienza – Dip. di Ingegneria Astronautica, Elettrica ed Energetica - DIAEE e Politecnico di Torino, Dip. Energia - DENERG), e di aziende private per la fornitura di consulenze e dei servizi di cui sopra:

- **LA1.2:** introduzione di funzioni statistiche per calcolo di prestazioni/costi di tecnologie a larga diffusione (residenziale, trasporti) alimentato da dati di mercato;
- **LA1.3:** introduzione di un nuovo DB su misure di incentivazione per le tecnologie rinnovabili e l'efficienza energetica;
- **LA1.5-1.6:** prove di applicazione e successiva implementazione/addestramento di SW basati su tecnologia AI per lettura automatica documenti e gestione dati, reporting, virtual assistant (in collaborazione con fornitore di SW-AI);

- **LA1.7:** ammodernamento HW del portale SiMTE (server, etc.);
- **LA1.9-1.12:** aggiornamento e produzione ex-novo di documenti e dati su tecnologie energetiche (settori: produzione, trasmissione e distribuzione di elettricità e calore; residenziale e servizi) in collaborazione con Politecnico di Torino e Università di Roma La Sapienza (circa 50 documenti/brief);
- **LA1.18-1.19:** valutazioni energetiche regionali: implementazione della metodologia sul portale SiMTE e applicazione alla Regione Basilicata;
- **LA1.22:** Predisposizione di una strategia di diffusione del portale tramite contratto assegnato ad una azienda di settore.

La terza annualità 2021 comprende 7 LA e prevede il completamento delle attività progettuali:

- **LA1.4:** applicazione dei modelli di calcolo prestazione e costi alle nuove tecnologie inserite nella banca dati nel 2020-21 (LA1.9-1.16) e adattamento dei modelli ai nuovi SW-AI (LA1.5-1.6);
- **LA1.13-1.16:** aggiornamento e produzione ex-novo di documenti e dati su tecnologie (settori: produzione e trasformazione di energia primaria; industria; trasporti) con Politecnico di Torino e Università di Roma La Sapienza (circa 45 doc./brief);
- **LA1.20:** Valutazioni energetiche regionali, applicazione alla Regione Piemonte;
- **LA1.23:** attuazione della strategia diffusione: lancio del portale SiMTE, organizzazione di eventi, workshop, etc.) in collaborazione con azienda di settore.

Il costo complessivo atteso del progetto è di € 2.316.797,94. La Figura 1 illustra la ripartizione del budget per voce di spesa (incluso spese generali), per anno, per obiettivo e per beneficiario (escluso spese generali). La Figura 2 illustra la ripartizione dei costi per voce di spesa nel triennio (escluso spese generali).

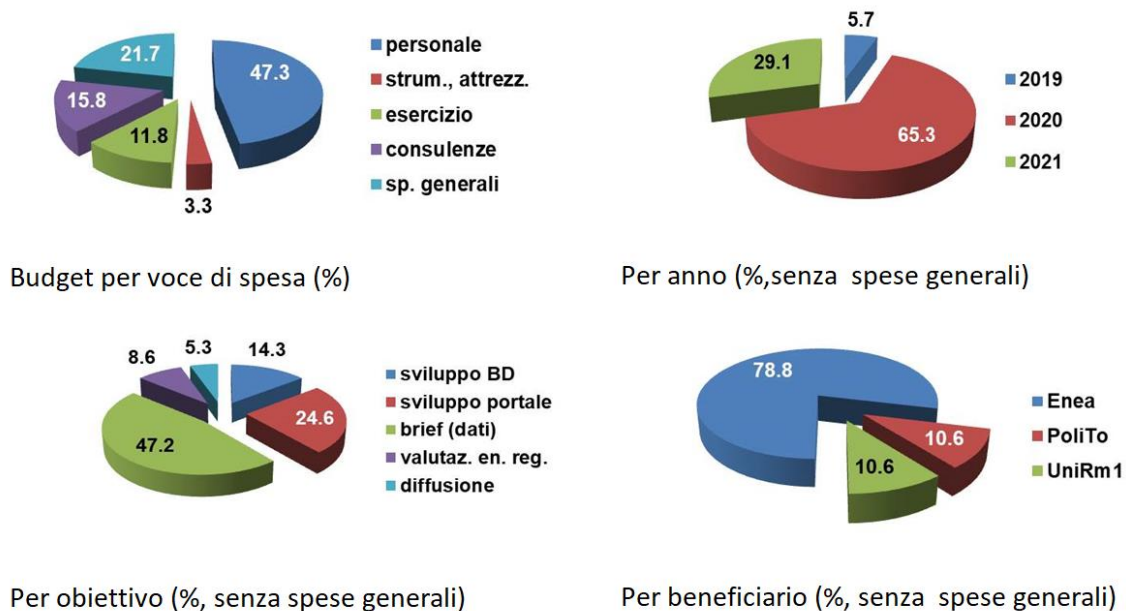


Figura 1 – Ripartizione del budget per voce di spesa, per anno, per obiettivo e per beneficiario.

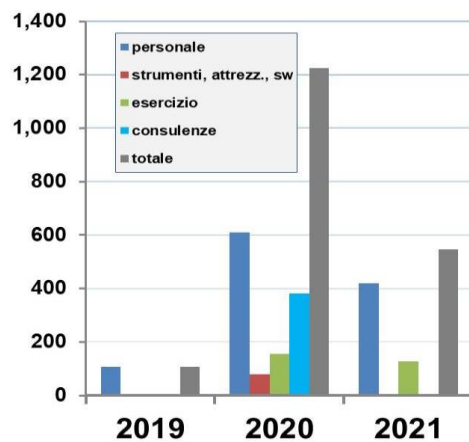


Figura 2 – Costi del progetto (k€) per voce di spesa nel triennio 2019-21 (senza spese generali)

1.1 Impatto della pandemia Covid-19 sulle attività previste per l'anno 2020

La pandemia Covid-19 che ha investito il paese all'inizio del 2020 ha causato ritardi nell'approvazione formale del progetto (Giugno 2020) e un conseguente ritardo nella disponibilità dei fondi. Successivamente, la riorganizzazione delle attività lavorative conseguente alla pandemia, unitamente ai tempi tecnici richiesti dalle procedure di gara per l'acquisto di forniture e consulenze esterne, non hanno consentito l'avvio e il regolare svolgimento di alcune linee di attività previste per il 2020. In particolare, non è stato possibile avviare la **LA 1.5** (Sviluppo portale: Valutazione e prova di applicabilità di strumenti SW basati su tecnologie AI), la **LA 1.6** (Sviluppo portale: Implementazione e addestramento di SW basati su AI). Conseguentemente, non è stato possibile avviare la **LA 1.7** (Sviluppo portale: ammodernamento HW), **LA 1.2** (Sviluppo data-base: Implementazione di funzioni statistiche per analisi prestazioni e costi) e la **LA 1.22** (Attività di diffusione 2020), queste ultime dipendenti per diverse ragioni dalle precedenti LA 1.5 e 1.6 e ad esse collegate. In particolare, la gara per la fornitura di SW-AI e per gli associati servizi di prova, implementazione e addestramento del SW stesso è stata aggiudicata in via provvisoria soltanto in data 16 Dicembre 2020 e le attività relative alle LA 1.5 e 1.6 sono ufficialmente iniziate nell'Aprile 2021 (verbale inizio lavori) e comprese nei restanti 8 mesi del 2021 a fronte di una durata inizialmente prevista per le LA 1.5 e 1.6 di complessivi 12 mesi (dal Gennaio al Dicembre 2020).

A tale proposito va rilevato che le attività di implementazione/addestramento del SW-AI non possono essere comprese in tempi sensibilmente più brevi di quelli previsti in ragione del loro carattere fortemente innovativo che riguarda l'addestramento del SW-AI alla lettura automatica di documenti tecnici e pubblicazioni tecnico-scientifiche, con riconoscimento ed estrazione di dati tecnico-economici di interesse (prestazioni, emissioni e costi delle tecnologie per la produzione e uso finale dell'energia) nelle rispettive unità di misura, e successiva gestione e restituzione di dati e informazioni su richiesta dell'utente della piattaforma SIMTE. Le interazioni con gli esperti della società vincitrice della gara (Sistemi Informativi/IBM) hanno confermato che l'applicazione in questione è assolutamente innovativa e di frontiera.

Allo sviluppo delle LA 1.5 e 1.6 sono peraltro collegate per le ragioni di seguito esposte la LA 1.7 (l'ammodernamento HW del portale dipende dall'implementazione del SW-AI), la LA 1.2 (le funzioni di analisi statistica hanno applicazione in presenza di una larga base acquisibile solo con il contributo di SW-AI), e la LA 1.22 (la promozione verso l'utenza dei servizi offerti dalla piattaforma SIMTE è fortemente caratterizzata dalla presenza di SW-AI e dell'associato *Assistente Virtuale*, soprattutto per quanto riguarda la facilità di fruizione dei servizi offerti da parte dell'utenza).

Per le motivazioni sopra esposte è stata proposta ed accettata una variante di progetto che posticipa all'annualità 2021 lo svolgimento delle LA 1.5, 1.6, 1.7, 1.2 e 1.22.

Nel corso della terza annualità (periodo 01.01.2021 – 31.12.2021) sono state quindi svolte attività relative a 12 linee di attività ed in particolare **LA1.2, LA1.4, LA1.5, LA1.6, LA1.7, LA1.13, LA1.14, LA1.15, LA 1.16, LA1.20, LA1.22 e LA1.23**. Di queste:

- le LA 1.2, 1.5, 1.6 e 1.7 riguardano l'aggiornamento SW e HW del portale e del database della piattaforma SiMTE con introduzione di strumenti basati su intelligenza artificiale AI;
- le LA1.13-16 riguardano l'aggiornamento e/o la redazione ex novo di circa 50 documenti tecnici (relativi ad altrettante tecnologie energetiche) della piattaforma SiMTE. Si tratta di una attività divisa in 4 diverse LA per ragioni contabili e di rendicontazione in quanto affidata, oltre che ad Enea, a due diversi Co-beneficiari Universitari (Politecnico di Torino e Sapienza di Roma) e relativa a tre diversi macro-settori delle tecnologie energetiche (produzione e trasformazione dell'energia primaria, industria e trasporti);
- la LA 1.20 riguarda la valutazione energetica della regione Piemonte;
- le LA 1.22 e 1.23 riguardano le attività di diffusione

Le 12 LA di cui sopra hanno impegnato complessivamente risorse di personale per circa ore 20950 ore corrispondenti ad un impegno complessivo di circa **12.25 persona-anno** di personale professionale di diversi livelli di esperienza.

1.2 Brevi informazioni sulla piattaforma SiMTE

La banca dati e il portale SiMTE sono stati realizzati da Enea nel periodo Febbraio 2016-Gennaio 2018 per conto del MiSE. Si tratta, come previsto dai già citati dispositivi di legge, di un sistema informativo e di monitoraggio dello stato e delle prospettive delle tecnologie di offerta e domanda dell'energia a supporto di operatori e decisori pubblici e privati del settore energetico. Il sistema può essere inoltre di ausilio per affrontare le scelte di politica energetica necessarie per realizzare la rapida transizione energetica ed ecologica che dovrà caratterizzare il decennio 2020-30 al fine di far fronte alla questione climatica.

Nella configurazione 2018 il portale SiMTE (www.simte.enea.it, Figura 3) era aperto in via *sperimentale* in attesa del necessario aggiornamento ed ampliamento. La banca dati offriva infatti documenti, informazioni e dati (prestazioni tecniche, consumi energetici, emissioni, normative, quote di mercato, peculiarità nazionali, costi di investimento, di esercizio, di prodotto/servizio, etc.) sulle tecnologie energetiche aggiornati al Gennaio 2018 e conteneva circa 50 tecnologie o cluster tecnologici divisi nei 5 settori già elencati in precedenza (produzione e trasformazione di energia primaria; produzione e distribuzione di elettricità e calore; residenziale e servizi; trasporti; industria). Al termine dell'attuale progetto è previsto il raddoppio delle tecnologie presenti sulla piattaforma.

Per ogni tecnologia o cluster di tecnologie (Figura 4) i dati e le informazioni vengono discussi e contestualizzati in un documento di base (brief) accessibile all'utenza, redatto da esperti e periodicamente aggiornato. I dati, trasferiti in banca dati, sono accessibili all'utenza mediante sistemi di ricerca e analisi che consentono anche il confronto tra varie tecnologie a richiesta dell'utente. I dati vengono poi essere trasferiti ad un modello che ricalcola prestazioni e costi delle tecnologie in base a criteri omogenei. L'input del modello, derivato inizialmente dai dati presenti in banca-dati e reso pienamente trasparente per l'utente, può essere modificato con dati immessi dall'utente stesso. Il modello è applicabile a tutte le tecnologie di domanda/offerta. Di conseguenza, pur non sostituendosi alle accurate analisi costi-benefici a cura dell'investitore, il modello consente confronti costi/benefici (prestazioni, emissioni, costi) tra tecnologie anche afferenti a settori diversi. Esso è destinato prevalentemente ad utenti "esperti" (con conoscenze nel settore dell'analisi costi-benefici delle tecnologie energetiche) e consente anche l'analisi di

sistemi complessi (es: efficientamento nel settore residenziale, cogenerazione, impianti con *energy storage*). In tal caso l'utente potrà anche avvalersi di casi pilota già elaborati, disponibili sul portale.

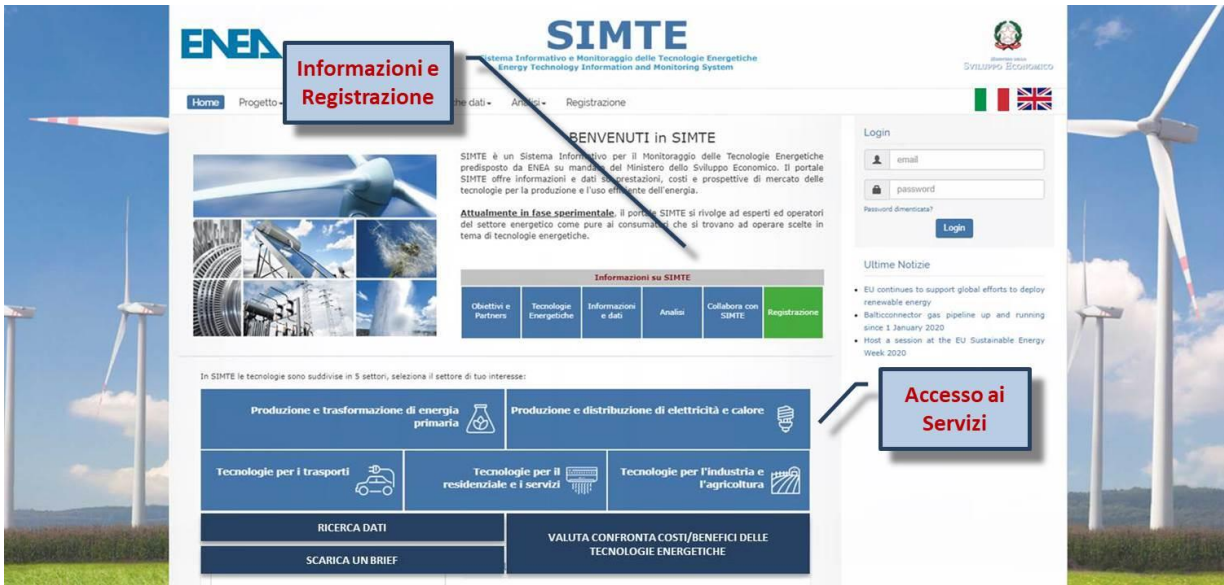


Figura 3 – Homepage della piattaforma SIMTE (www.simte.enea.it oppure google “simte”)

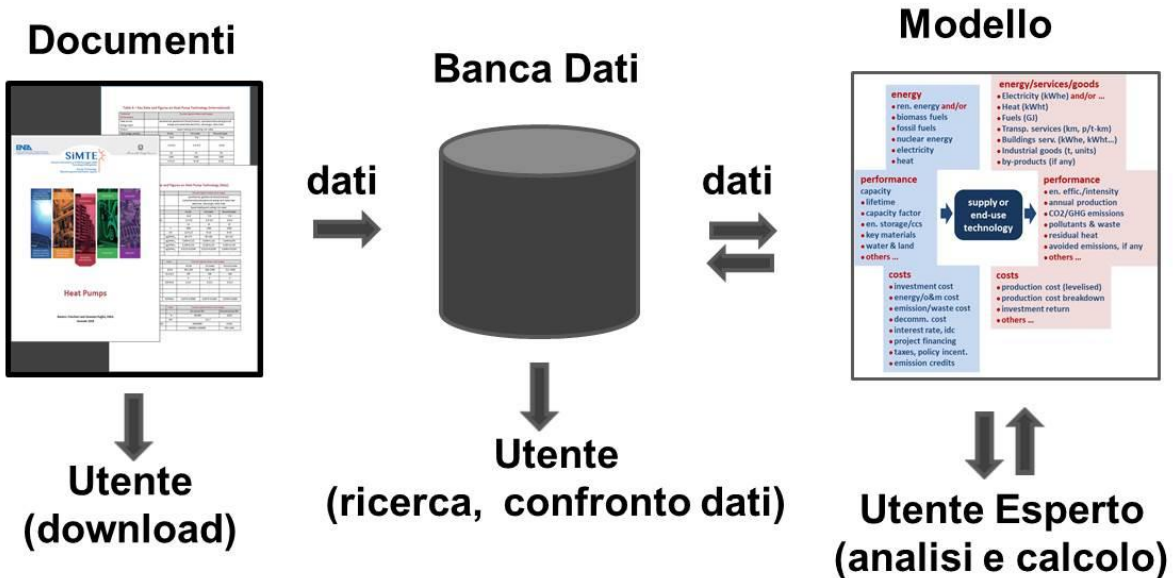


Figura 4 – Struttura, flusso dati e servizi della Banca-dati SIMTE

2 Descrizione delle attività svolte e risultati della LA 1.6

2.1 Introduzione

Come previsto dal Piano Triennale Realizzativo (PTR) del progetto, nel corso dell'annualità 2021 sono stati svolti ed ultimati i lavori relativi alla linea di attività LA 1.6 che riguarda l'implementazione sulla piattaforma SiMTE di un software (SW) cognitivo basato su intelligenza artificiale (AI, Artificial Intelligence) per svolgere le seguenti funzioni:

a) lettura automatica di documenti tecnico-economici sulle tecnologie energetiche con riconoscimento ed estrazione dei dati di interesse (prestazioni, costi, aspetti di mercato) e trasferimento dei dati alla banca dati SiMTE;

b) assistenza agli utenti della piattaforma SiMTE (mediante un assistente virtuale) nell'utilizzo dei servizi offerti tra i quali :

- dati e informazioni sulle tecnologie energetiche (accesso alla banca dati), ivi incluso il download dei documenti presenti sulla piattaforma stessa;
- modello di calcolo per l'analisi tecnico-economica delle tecnologie (analisi costi-benefici, con dati presenti nella banca dati o immessi dall'utente).

L'introduzione dell'AI sulla piattaforma SiMTE costituisce la parte più **innovativa e sperimentale** del progetto. Essa è motivata dalla necessità di ridurre i notevoli oneri (in termini di tempo, risorse umane e costi) derivanti dal necessario continuo aggiornamento della piattaforma (per tenere il passo con l'evoluzione di innovazioni, prestazioni, costi e aspetti di mercato delle tecnologie energetiche) e di rendere più fruibili per l'utenza i servizi offerti. In assenza di SW cognitivo, l'aggiornamento dei dati e dei documenti presenti sulla piattaforma, come pure l'acquisizione di nuovi documenti e il trasferimento dei dati alla banca dati, sono processi che impegnano esperti dei vari settori tecnologici per tempi considerevoli. Inoltre, attraverso un *assistente virtuale*, l'AI è in grado di semplificare per l'utenza la fruizione dei servizi offerti dalla piattaforma guidando l'utente nelle varie operazioni per l'accesso alla banca dati e al modello.

L'implementazione di SW cognitivo richiede tuttavia non solo la fornitura del SW ma un vero e proprio processo di addestramento del SW stesso per le funzioni da svolgere, che richiede competenze specialistiche. Per tale ragione l'implementazione del SW ha richiesto l'assegnazione (tramite gara di appalto) di un contratto non solo per la fornitura (accesso al cloud ove il SW risiede) ma anche per l'addestramento del SW per svolgere le funzioni previste, oltre ad un training del personale ENEA per acquisire autonomia nella gestione e nell'utilizzo del SW (ivi compreso le competenze per l'addestramento).

La natura innovativa dell'applicazione e la ridotta conoscenza delle reali potenzialità dei nuovi SW cognitivi in relazione agli obiettivi preposti hanno indotto a dividere l'attività di implementazione in due fasi distinte e in due diverse linee di attività:

- LA 1.5 - Verifica di applicabilità dell'AI alle esigenze della piattaforma SiMTE (Fase 1)
- LA 1.6 - Implementazione dell'AI sulla piattaforma SiMTE (Fase 2, a valle di esito positivo della Fase 1).

La prima fase è stata dedicata in primo luogo allo scambio di competenze tra committente e fornitore (rispettivamente, tecnologie energetiche e tecnica di addestramento) al fine di pervenire ad una reciproca comprensione di esigenze ed obiettivi e, in secondo luogo, alla verifica della capacità del SW di raggiungere gli obiettivi prefissati;

La seconda fase ha riguardato l'effettiva implementazione del SW, l'addestramento e la messa a punto di tutte le funzionalità previste.

Una terza linea di attività, la LA 1.2, che prevede la implementazione sulla piattaforma SiMTE di funzioni statistiche per analisi di prestazioni e costi di tecnologie energetiche di larga diffusione commerciale (es: veicoli, elettrodomestici) è confluita in corso d'opera nella implementazione dell'AI poiché le funzioni statistiche, con un SW ed un lavoro aggiuntivo, sono state integrate nel sistema guidato dall'AI.

In ottemperanza alle regole di *reporting* della Ricerca di Sistema Elettrico, le tre LA di cui sopra (1.2, 1.5 e 1.6) sono oggetto di rapporti tecnici distinti (rispettivamente RdS/PTR2021/068, 070, 071) ma è evidente che sul piano sostanziale le LA 1.2 e 1.5 sono strettamente legate alla LA 1.6 e i rispettivi rapporti condividono parti del presente rapporto.

Vale la pena infine di ricordare che l'implementazione dell'AI era inizialmente prevista per l'annualità 2020 per un tempo di espletamento complessivo di 12 mesi (4 mesi per la Fase 1 e 8 mesi per la Fase 2). I ritardi conseguenti all'emergenza Covid unitamente ad alcune difficoltà dell'Amministrazione ENEA non hanno consentito tuttavia di procedere celermente all'assegnazione di un contratto di fornitura nell'ambito della cosiddetta "Convenzione CRUI", ambito nel quale, a suo tempo è stata espletata una gara internazionale (vinta da IBM) per la fornitura di SW-AI ad Università ed enti di ricerca. E' stato quindi necessario istruire una nuova gara di appalto internazionale i cui tempi di espletamento hanno portato ad iniziare i lavori soltanto nell'Aprile 2021 e alla necessità di comprimerli in poco più di 6 mesi. Oltre alla notevole compressione dei tempi di lavoro, è stato anche necessario richiedere l'approvazione di una variante di progetto per spostare al 2021 tutte le LA collegate alla introduzione dell'AI.

Nel seguito vengono descritti in sintesi le varie fasi del lavoro di implementazione e addestramento dell'AI e i risultati ottenuti, rimandando agli allegati tecnici (ALL 1 e ALL2) redatti dalla società vincitrice della gara (Sistemi Informativi, gruppo IBM) per maggiori dettagli tecnici.

2.2 Attività svolte nella Fase 1

Come accennato nella introduzione, al fine di pervenire ad maggiore comprensione delle potenzialità (e dei limiti) del SW cognitivo da parte del Committente (ENEA) e degli obiettivi da raggiungere da parte del Fornitore del SW (Sistemi Informativi/IBM), nella parte iniziale della Fase 1 sono state messe in comune le conoscenze e la terminologia informatica del Fornitore con quelle del Committente in tema di caratterizzazione tecnico-economica delle tecnologie energetiche. Il Fornitore ha dovuto familiarizzare con termini quali *efficienza e intensità energetica, fattore di capacità, vita operativa, costi di investimento e di esercizio delle tecnologie* (e con le rispettive unità di misura) mentre il Committente ha dovuto familiarizzare con il significato attribuito a termini quali *intenti, entità e relazioni, indicatori*, utilizzati nella descrizione delle tecniche di addestramento dei SW cognitivi. Solo al termine di questa fase di *merging* delle competenze è stato possibile per il Fornitore comprendere con il necessario dettaglio gli obiettivi da raggiungere e le caratteristiche del sistema da realizzare e, per il Committente, comprendere le reali potenzialità e i limiti degli attuali SW cognitivi.

La fase di *merging* delle competenze ha evidenziato, tra l'altro, come l'assegnazione (tramite gara) di un contratto per la fornitura di servizi (imposta in base alla generica interpretazione della normativa vigente) mal si presti a realizzare una applicazione sperimentale del tipo in questione. Le procedure di gara non consentono infatti un approfondito dialogo preliminare tra i partecipanti alla gara e il committente per arrivare ad una precisa definizione di obiettivi raggiungibili. In pratica il committente non può conoscere in misura sufficiente i limiti e le potenzialità del SW cognitivo e i fornitori non possono conoscere in misura sufficiente l'ambito operativo e le esigenze del committente. Meglio sarebbe procedere con l'assegnazione diretta di un contratto di ricerca dopo una maggiore comprensione delle reciproche esigenze.

Al termine della fase di *merging* è stato quindi possibile iniziare il lavoro con la cosiddetta "**Raccolta dei requisiti funzionali**" (vedi Allegato 2).

Parallelamente alla raccolta requisiti funzionali, è stato preparato l'ambiente informatico operativo. Il SW cognitivo risiede infatti fisicamente su IBM Cloud. Sono quindi stati stabiliti i necessari collegamenti

informatici tra la piattaforma SiMTE residente su server ENEA e il Cloud IBM, inizializzando i servizi necessari per lo sviluppo del sistema e replicando sul Cloud la struttura del database SiMTE per il salvataggio dei dati estratti dai documenti.

Nella seconda parte della Fase 1 è stata verificata l'applicabilità del SW per gli obiettivi fissati mediante una applicazione dimostrativa ad alcune **tecnologie pilota**. In questo ambito sono state fatte alcune scelte relative alla lingua di addestramento del SW e alle tecnologie pilota e si è proceduto all'addestramento sulle tecnologie pilota per poi verificarne le funzionalità.

Per velocizzare il processo (dati i ristretti tempi a disposizione rispetto alla programmazione iniziale), è stato deciso di procedere all'addestramento del SW in lingua inglese per una duplice ragione: pur potendo operare in molte lingue, il SW è nativo in lingua inglese; inoltre la terminologia utilizzata nel settore delle tecnologie energetiche è largamente diffusa e utilizzata in lingua inglese. L'addestramento in lingua inglese è stato possibile poiché tutti i documenti sulle tecnologie energetiche presenti sulla piattaforma SiMTE sono stati prodotti in doppia lingua inglese-italiano, nella prospettiva di rendere accessibile il sistema anche in ambito internazionale.

Per quanto riguarda la scelta delle tecnologie pilota, sono state selezionate alcune delle più diffuse tecnologie del settore della elettro-generazione, in particolare, produzione di elettricità da combustibili fossili (COAL, coal-fired power), da solare fotovoltaico (PV, solar photovoltaics) e da solare a concentrazione (CSP, concentrating solar power) poiché la terminologia tecnico-economica utilizzata in tale settore è più largamente diffusa e standardizzata (es: *energy efficiency, capacity factor, lifetime, overnight investment cost, levelized cost of electricity, LCOE*).

Il SW utilizzato per l'applicazione è il sistema *IBM Watson* costituito da una famiglia di diversi SW cognitivi, ognuno dei quali preposto ad una specifica funzione (*Watson Assistant, Watson Discovery, Watson Knowledge Studio, Watson Natural Language Understanding*), coordinati da un SW cosiddetto "*Orchestratore*" che, in base ai requisiti funzionali e al progetto dell'architettura informatica del sistema, permette di integrare le funzioni dei vari componenti. Una descrizione delle funzioni dei vari SW, delle fasi di addestramento del SW cognitivo e della terminologia utilizzata (es: intenti, entità, relazioni) e viene fornita nel seguito (per descrizioni più dettagliate e in linguaggio più tecnico si rimanda all'Allegato 1 per i componenti del SW e all'Allegato 2 per le fasi di addestramento).

Le attività svolte in questa fase possono quindi essere riassunte come segue:

- analisi dei requisiti funzionali e definizione dell'architettura del sistema;
- implementazione del nucleo del SW cognitivo (orchestratore e componenti);
- integrazione con la base dati del sistema SiMTE, replicato su cloud IBM;
- sviluppo e integrazione dell'interfaccia per la funzione di **Assistente Virtuale**;
- *ingestion* dei documenti SiMTE sottoposti al SW cognitivo e implementazione delle funzioni di storage e download dei documenti.

Nella fase iniziale alcune sessioni di lavoro congiunto tra esperti ENEA e IBM/Sistemi sono state dedicate alla definizione di una sorta di **glossari** di termini (categorie, parametri indicatori, grandezze fisiche, tecniche ed economiche, con rispettivi sinonimi, varianti semantiche e unità di misura) che intervengono nella caratterizzazione tecnico-economica delle tecnologie energetiche. Tali glossari sono stati poi utilizzati nella varie fasi di addestramento del SW unitamente ai documenti relativi alle tecnologie pilota selezionate.

Di seguito vengono descritte le principali fasi di preparazione del sistema e di addestramento del SW cognitivo.

Assistente Virtuale

L'**Assistente Virtuale** è un componente fondamentale del sistema in quanto presiede a tutte le interazioni con l'utente della piattaforma SiMTE. Per lo sviluppo della funzione di Assistente Virtuale sono stati

sviluppati i componenti grafici di base dell'interfaccia utente (Figura 5) e i necessari collegamenti con l'Orchestratore e con i componenti Watson Assistant e Watson Discovery.

Il SW Watson Assistant consente di aggiungere alle applicazioni sviluppate la funzionalità di conversazione con l'utente in linguaggio naturale. È stato quindi creato l'assistente *ENEA AI SiMTE*, dotato di due "skills", corrispondenti a differenti capacità: una capacità di *Search*, in grado di richiamare le informazioni dal SW Watson Discovery che presiede alla gestione dei documenti, ed una capacità di *Dialog*, per permettere all'Assistente di comprendere e rispondere a domande e richieste dell'utente utilizzando tecniche di machine learning ed elaborazioni in linguaggio naturale. Le attività di addestramento dell'**Assistente Virtuale** possono essere suddivise in tre categorie relative agli **intenti**, alle **entità**, e al **dialogo**.

Per creare le conversazioni si parte dagli **intenti** che servono a categorizzare l'input dell'utente, "incanalando" le diverse tipologie di richieste. Per ogni intento l'assistente è stato addestrato con frasi di esempio per consentire l'identificazione dell'input dell'utente da parte dell'Assistente. Successivamente alla creazione e categorizzazione degli intenti, sono state create le **entità**, che rappresentano termini rilevanti per gli intenti gestiti e che forniscono quindi un contesto distintivo per un dato intento. In altri termini, le entità sono valori specifici (in questo caso, parole) da cercare nell'input dell'utente, con vari sinonimi per ciascun valore; esse rappresentano gli oggetti di interesse che aiutano a caratterizzare l'input (e quindi l'intenzione) dell'utente. Dopo aver definito intenti ed entità, è stato costruito il **dialogo**. Tale costruzione è stata fatta in maniera grafica (Figura 6), attraverso una serie di nodi collegati tra loro. Ogni nodo contiene condizioni che vengono valutate, al fine di accedere o meno ai relativi output. La valutazione dei nodi avviene dall'alto verso il basso.

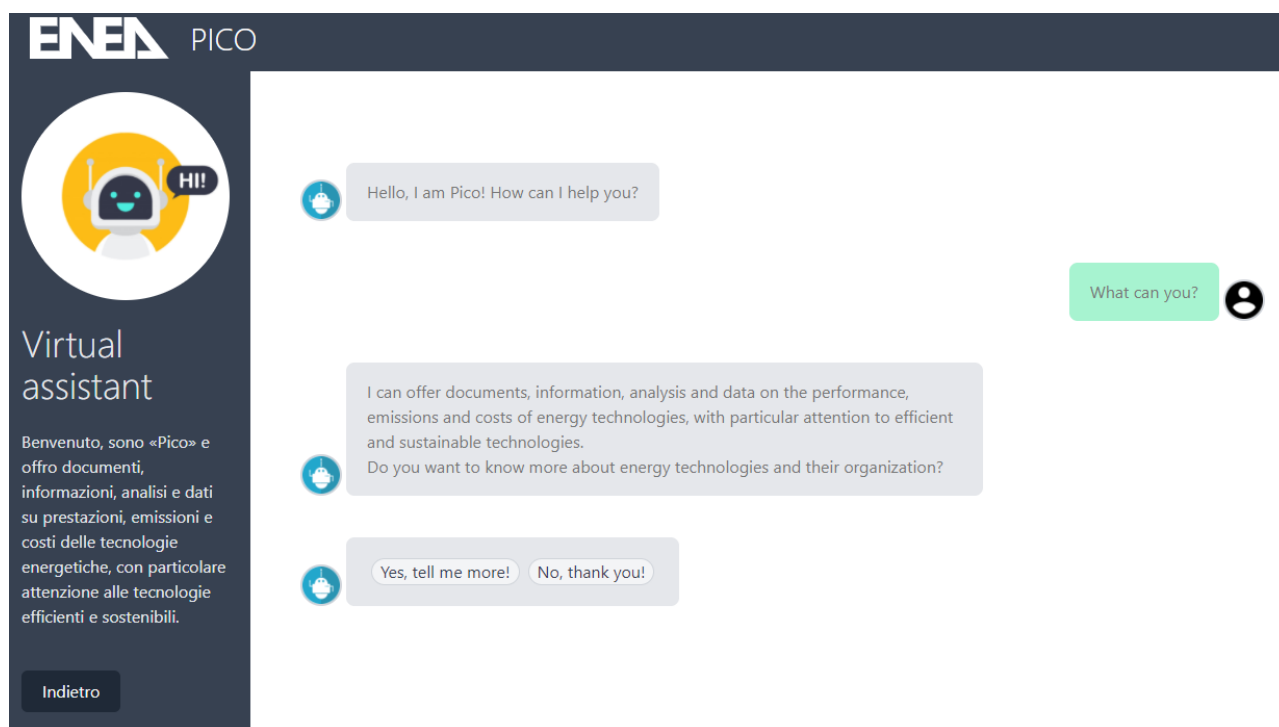


Figura 5 - Interfaccia Grafica dell'Assistente Virtuale "Pico"

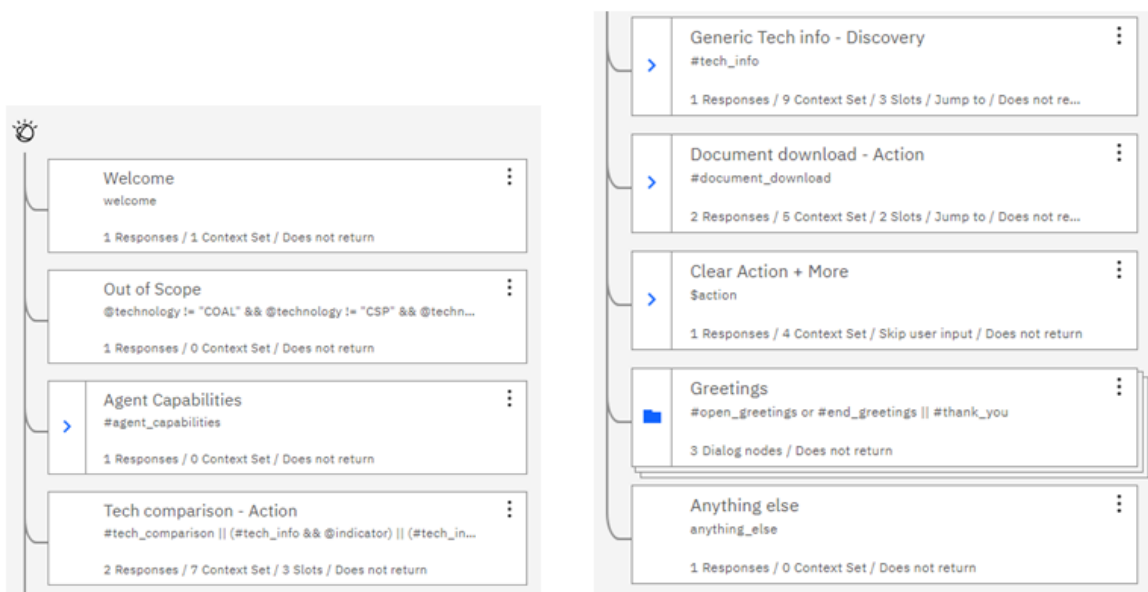


Figura 6 - Dialogo dell'Assistente Virtuale

Gestione dei documenti

Le attività relative alla gestione dei documenti coinvolgono i componenti Watson Knowledge Studio e Watson Discovery.

Il componente Watson Knowledge Studio genera modelli cognitivi personalizzati che identificano ed estraggono, da testi non strutturati, **entità e relazioni** univoche per il settore o dominio su cui il modello viene addestrato. In pratica il componente Watson Knowledge Studio presiede un flusso di lavoro (workflow) costituito da una serie di passi del tipo rappresentato in Figura 7.

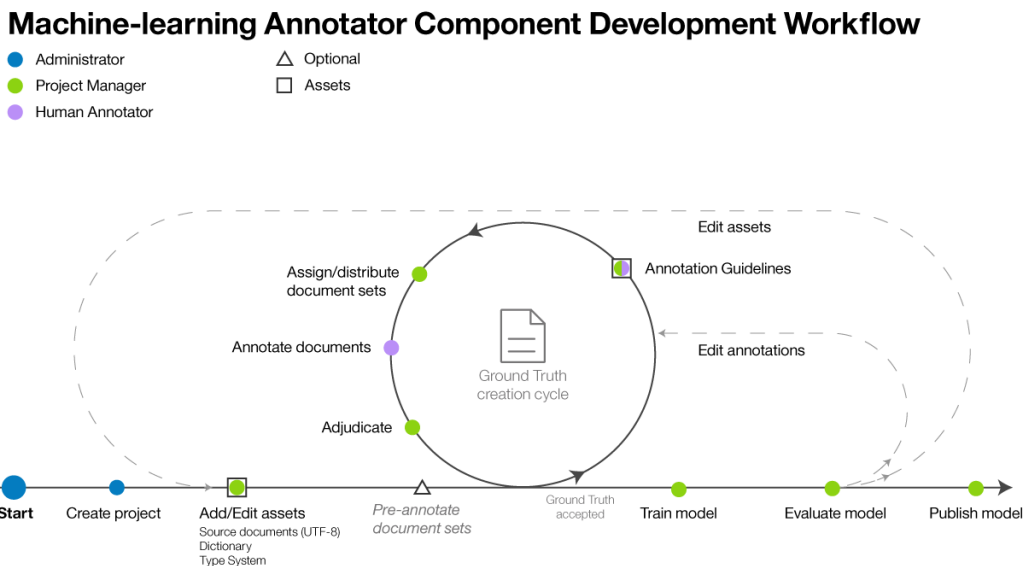


Figura 7 - Workflow di Watson Knowledge Studio per la creazione di un modello Machine Learning.

Il primo passo del workflow prevede la creazione del workspace dedicato al progetto, da generare nella lingua dei documenti che dovranno poi essere analizzati dal modello. Per il progetto SiMTE è stato creato il workspace "ENEA AI-SiMTE" in lingua inglese.

Il secondo passo del workflow prevede l'aggiunta degli asset, che si dividono in **type system**, **dizionari**, e **documenti di origine**.

Il **type system** (sistema tipo) definisce le **entità** e le **relazioni** tra le entità che si ritiene debbano essere estratte dai documenti. In particolare, una entità è una categorizzazione di una cosa reale, una "etichetta" da associare ad un particolare concetto; una relazione è un'associazione tra due entità, quindi un collegamento tra due concetti.

Il type system utilizzato in Fase 1 è stato definito in una serie di incontri tra Committente e Fornitore nell'ambito dei quali sono stati tra l'altro definiti i "**glossari**" precedentemente menzionati. Per la Fase 1 sono state aggiunte **106 entità e 5 "macro relazioni"**, che collegano fra loro più entità, per un totale di **1.218 relazioni effettive** (di seguito vengono riportati gli elenchi delle 106 entità utilizzate e delle 5 relazioni con relative descrizioni).

I **dizionari** raggruppano parole e/o frasi che devono essere trattate allo stesso modo dal modello, quindi parole o frasi che, pur non avendo uguale significato, identificano la stessa entità o la stessa relazione. Per la Fase 1 del progetto, sono stati aggiunti dizionari utili per la definizione delle entità, partendo dai termini stabiliti in fase di analisi, integrati con ulteriori definizioni estratte dai documenti oggetto di Fase 1.

I **documenti di origine** (source document) sono i documenti utilizzati per costruire il modello, per l'addestramento e per testare successivamente le funzionalità del modello stesso. Per la Fase 1 del progetto sono stati aggiunti otto documenti (tre per la tecnologia PV, tre per la tecnologia CSP, e due per la tecnologia COAL) suddivisi in otto set di training, e tre set di test. I set di training sono stati poi utilizzati per effettuare il training del modello, mentre quelli di test sono stati usati per testare le funzionalità del modello stesso.

Entità	Descrizione
CONVERSION_EFFICIENCY	L'entità fa riferimento all'indicatore <i>Annual Conversion Efficiency</i> .
AVAILABILITY_FACTOR	L'entità fa riferimento all'indicatore <i>Availability Factor</i> .
AVOIDED_CO2_EMISSIONS	L'entità fa riferimento all'indicatore <i>Avoided CO₂ Emissions</i> .
BOS_COST	L'entità fa riferimento all'indicatore <i>Balance Of System Cost</i> .
CAPACITY	L'entità fa riferimento all'indicatore <i>Capacity</i> .
CAPACITY_FACTOR	L'entità fa riferimento all'indicatore <i>Capacity Factor</i> .
CAPACITY_UNDER_CONSTRUCTION	L'entità fa riferimento all'indicatore <i>Capacity Under Construction</i> .
CELL_EFFICIENCY	L'entità fa riferimento all'indicatore <i>Cell Efficiency</i> .
CO2_EMISSIONS	L'entità fa riferimento all'indicatore <i>CO₂ Emissions</i> .
COAL	L'entità fa riferimento alla tecnologia del <i>Carbone (COAL)</i> .
COAL_IGCC	L'entità fa riferimento alla variante tecnologica <i>Integrated Gasification Combined Cycle (IGCC) Technology</i> del COAL.
COAL_PC	L'entità fa riferimento alla variante tecnologica <i>Pulverized Coal Technology</i> del COAL.
COAL_PC_AUSC	L'entità fa riferimento alla sotto variante tecnologica della <i>Pulverized Coal Technology</i> , la <i>Advanced Ultra-Super-Critical Cycle (AUSC) Technology</i> .
COAL_PC_SC	L'entità fa riferimento alla sotto variante tecnologica della <i>Pulverized Coal Technology</i> , la <i>Super-Critical Cycle (SC) Technology</i> .
COAL_PC_SUB	L'entità fa riferimento alla sotto variante tecnologica della <i>Pulverized Coal Technology</i> , la <i>Sub-Critical Cycle Technology</i> .
COAL_PC_USC	L'entità fa riferimento alla sotto variante tecnologica della <i>Pulverized Coal Technology</i> , la <i>Ultra-Super-Critical Cycle (USC) Technology</i> .
COMMERCIAL_MODULE_EFFICIENCY	L'entità fa riferimento all'indicatore <i>Commercial Module Efficiency</i> .
CONCENTRATION_RATIO	L'entità fa riferimento all'indicatore <i>Concentration Ratio</i> .
CONSTRUCTION_TIME	L'entità fa riferimento all'indicatore <i>Construction Time</i> .
COST_BREAKDOWN	L'entità fa riferimento all'indicatore <i>Cost Breakdown</i> .
CSP	L'entità fa riferimento alla tecnologia del <i>Concentrating Solar Power (CSP, o Solar Thermal Electricity)</i> .
CSP_VAR_CR	L'entità fa riferimento alla variante tecnologica <i>Central Receiver Technology</i> del CSP.
CSP_VAR_CR_SALT_WS	L'entità fa riferimento alla sotto variante tecnologica della <i>Linear Fresnel Reflector Technology</i> , la <i>Tower HTF Salt With Storage Technology</i> .

CSP_VAR_CR_STEAM_NS	L'entità fa riferimento alla sotto variante tecnologica della <i>Linear Fresnel Reflector Technology</i> , la <i>Tower HTF Water Without Storage Technology</i> .
CSP_VAR_DISH	L'entità fa riferimento alla variante tecnologica <i>Dish System Technology</i> del CSP.
CSP_VAR_LFR	L'entità fa riferimento alla variante tecnologica <i>Parabolic Trough Technology</i> del CSP.
CSP_VAR_LFR_STEAM_NS	L'entità fa riferimento alla sotto variante tecnologica della <i>Linear Fresnel Reflector Technology</i> , la <i>Fresnel HTF Water Without Storage Technology</i> .
CSP_VAR_LFR_STEAM_WS	L'entità fa riferimento alla sotto variante tecnologica della <i>Linear Fresnel Reflector Technology</i> , la <i>Fresnel HTF Water With Storage Technology</i> .
CSP_VAR_PT	L'entità fa riferimento alla variante tecnologica <i>Linear Fresnel Reflector Technology</i> del CSP.
CSP_VAR_PT_OIL_NS	L'entità fa riferimento alla sotto variante tecnologica della <i>Linear Fresnel Reflector Technology</i> , la <i>Trough HTF Oil Without Storage Technology</i> .
CSP_VAR_PT_OIL_WS	L'entità fa riferimento alla sotto variante tecnologica della <i>Linear Fresnel Reflector Technology</i> , la <i>Trough HTF Oil With Storage Technology</i> .
CSP_VAR_PT_OIL_WWS	L'entità fa riferimento alla sotto variante tecnologica della <i>Linear Fresnel Reflector Technology</i> , la <i>Trough HTF Oil With/Without Storage Technology</i> .
CSP_VAR_PT_SALT_WS	L'entità fa riferimento alla sotto variante tecnologica della <i>Linear Fresnel Reflector Technology</i> , la <i>Trough HTF Salt With Storage Technology</i> .
CSP_VAR_PT_STEAM_NS	L'entità fa riferimento alla sotto variante tecnologica della <i>Linear Fresnel Reflector Technology</i> , la <i>Trough HTF Water Without Storage Technology</i> .
CYCLE_PRESSURE	L'entità fa riferimento all'indicatore <i>Cycle Pressure</i> .
CYCLE_TEMPERATURE	L'entità fa riferimento all'indicatore <i>Cycle Temperature</i> .
CYCLE_TYPE	L'entità fa riferimento all'indicatore <i>Cycle Type</i> .
DECOMMISSIONING_COST	L'entità fa riferimento all'indicatore <i>Decommissioning Cost</i> .
ELECTRICITY_COST	L'entità fa riferimento all'indicatore <i>Electricity Cost</i> .
ELECTRICITY_SHARE	L'entità fa riferimento all'indicatore <i>Electricity Share</i> .
EMISSIONS	L'entità fa riferimento all'indicatore <i>Emissions</i> .
ENERGY_EFFICIENCY	L'entità fa riferimento all'indicatore <i>Energy Efficiency</i> .
ENERGY_INPUT_COST	L'entità fa riferimento all'indicatore <i>Energy Input Cost</i> .
ENERGY_PAYBACK_TIME	L'entità fa riferimento all'indicatore <i>Energy Payback Time</i> .
EXHAUSTED_HEAT_COOLING	L'entità fa riferimento all'indicatore <i>Exhausted Heat Cooling</i> .
EXTERNAL_COST	L'entità fa riferimento all'indicatore <i>External Cost</i> .
FINANCING_COST	L'entità fa riferimento all'indicatore <i>Financing Cost</i> .
FOM_COST	L'entità fa riferimento all'indicatore <i>Fixed Operation & Maintenance Cost</i> .
FUEL_COST	L'entità fa riferimento all'indicatore <i>Fuel Cost</i> .
GHG_EMISSIONS_COST	L'entità fa riferimento all'indicatore <i>GHG Emissions Cost</i> .
HEAT_STORAGE_MEDIA	L'entità fa riferimento all'indicatore <i>Heat Storage Media</i> .
HTF_TEMPERATURE	L'entità fa riferimento all'indicatore <i>HTF Temperature</i> .
HTF_TYPE	L'entità fa riferimento all'indicatore <i>HTF Type</i> .
INSTALLED_CAPACITY	L'entità fa riferimento all'indicatore <i>Installed Capacity</i> .
LAND_USE	L'entità fa riferimento all'indicatore <i>Land Use</i> .
LIFETIME	L'entità fa riferimento all'indicatore <i>Lifetime</i> .
MARKET_SHARE	L'entità fa riferimento all'indicatore <i>Market Share</i> .
MARKET_SHARE_PROJECTION	L'entità fa riferimento all'indicatore <i>Market Share Projection</i> .
MATERIAL_USE	L'entità fa riferimento all'indicatore <i>Material Use</i> .
MEASUREMENT_UNIT	L'entità fa riferimento all'unità di misura relativa ai valori associati agli indicatori.
MODULE_COST	L'entità fa riferimento all'indicatore <i>Module Cost</i> .
MODULE_EFFICIENCY	L'entità fa riferimento all'indicatore <i>Module Efficiency</i> .
NOMINAL_EFFICIENCY	L'entità fa riferimento all'indicatore <i>Nominal Efficiency</i> .
NOX_EMISSIONS	L'entità fa riferimento all'indicatore <i>NOx Emissions</i> .
NUMBER_OF_UNITS	L'entità fa riferimento all'indicatore <i>Number Of Units</i> .
OM_COST	L'entità fa riferimento all'indicatore <i>Operation & Maintenance Cost</i> .
PLANNED_CAPACITY	L'entità fa riferimento all'indicatore <i>Planned Capacity</i> .
PM_EMISSIONS	L'entità fa riferimento all'indicatore <i>PM Emissions</i> .
PRESSURE	L'entità fa riferimento all'indicatore <i>Pressure</i> .
PV	L'entità fa riferimento alla tecnologia del <i>Fotovoltaico (PV)</i> .
PV_APPLICATION_CS	L'entità fa riferimento al settore di applicazione del PV di tipo <i>Commercial</i> .
PV_APPLICATION_RS	L'entità fa riferimento al settore di applicazione del PV di tipo <i>Residential</i> .
PV_APPLICATION_US	L'entità fa riferimento al settore di applicazione del PV di tipo <i>Utility</i> .
PV_TYPE_BUILDING	L'entità fa riferimento all'installazione <i>Building Integrated PV System</i> , relativa alla tecnologiaPV.

PV_TYPE_GRID	L'entità fa riferimento all'installazione <i>Grid Connected PV System</i> , relativa alla tecnologia PV.
PV_TYPE_GROUND	L'entità fa riferimento all'installazione <i>Ground Based PV System</i> , relativa alla tecnologia PV.
PV_TYPE_OFFGRID	L'entità fa riferimento al tipo di installazione <i>Off-Grid PV System</i> , relativa alla tecnologia PV.
PV_TYPE_ROOF	L'entità fa riferimento al tipo di installazione <i>Rooftop PV System</i> , relativa alla tecnologia PV.
PV_VAR_CSI	L'entità fa riferimento alla variante tecnologica <i>Wafer-Based Crystalline Silicon Technology</i> PV.
PV_VAR_CSI_MC	L'entità fa riferimento alla sotto variante tecnologica del <i>Wafer-Based Crystalline Silicon Technology</i> , la <i>Single-Crystalline Silicon Technology</i> .
PV_VAR_CSI_SC	L'entità fa riferimento alla sotto variante tecnologica del <i>Wafer-Based Crystalline Silicon Technology</i> , la <i>Multi-Crystalline Silicon Technology</i> .
PV_VAR_NOVEL	L'entità fa riferimento alla variante tecnologica <i>Emerging And Novel PV Technology</i> del PV.
PV_VAR_NOVEL_AITF	L'entità fa riferimento alla sotto variante tecnologica della <i>Emerging And Novel PV Technology</i> , la <i>Advanced Inorganic Thin-Films Technology</i> .
PV_VAR_NOVEL_CPV	L'entità fa riferimento alla sotto variante tecnologica della <i>Emerging And Novel PV Technology</i> , la <i>Concentrating Photovoltaics Technology</i> .
PV_VAR_NOVEL_OSC	L'entità fa riferimento alla sotto variante tecnologica della <i>Emerging And Novel PV Technology</i> , la <i>Organic Solar Cells Technology</i> .
PV_VAR_NOVEL_OSC_DSSC	L'entità fa riferimento alla sotto variante tecnologica della <i>Organic Solar Cells Technology</i> della <i>Emerging And Novel PV Technology</i> , la <i>Hybrid Dye-sensitised Solar Cells Technology</i> .
PV_VAR_NOVEL_OSC_OPV	L'entità fa riferimento alla sotto variante tecnologica della <i>Organic Solar Cells Technology</i> della <i>Emerging And Novel PV Technology</i> , la <i>Organic Photovoltaic Cells Technology</i> .
PV_VAR_TF	L'entità fa riferimento alla variante tecnologica <i>Thin-Film Technology</i> del PV.
PV_VAR_TF_AS	L'entità fa riferimento alla sotto variante tecnologica del <i>Thin-Film Technology</i> , la <i>Amorphous Silicon Technology</i> .
PV_VAR_TF_CIGS	L'entità fa riferimento alla sotto variante tecnologica del <i>Thin-Film Technology</i> , la <i>Copper-indium-[gallium]-[di]selenide-[di]sulphide Film Technology</i> .
PV_VAR_TF_CT	L'entità fa riferimento alla sotto variante tecnologica del <i>Thin-Film Technology</i> , la <i>Cadmium-telluride Film Technology</i> .
PV_VAR_TF_MJS	L'entità fa riferimento alla sotto variante tecnologica del <i>Thin-Film Technology</i> , la <i>Multi-Junction Silicon Technology</i> .
SO2_EMISSIONS	L'entità fa riferimento all'indicatore <i>SO₂ Emissions</i> .
SOLAR_MULTIPLE	L'entità fa riferimento all'indicatore <i>Solar Multiple</i> .
STORAGE	L'entità fa riferimento all'indicatore <i>Storage</i> .
STORAGE_CAPACITY	L'entità fa riferimento all'indicatore <i>Storage Capacity</i> .
SYSTEM_COST	L'entità fa riferimento all'indicatore <i>System Cost</i> .
TEMPERATURE	L'entità fa riferimento all'indicatore <i>Temperature</i> .
TOTAL_CUMULATIVE_CAPACITY	L'entità fa riferimento all'indicatore <i>Total Cumulative Capacity</i> .
TYPICAL_SIZE	L'entità fa riferimento all'indicatore <i>Typical Size</i> .
UNITS_IN_OPERATION	L'entità fa riferimento all'indicatore <i>Units In Operation</i> .
VALUE	L'entità fa riferimento al valore numerico, tipicamente associato agli indicatori.
VALUE_MAX	L'entità fa riferimento al massimo valore numerico, tipicamente associato agli indicatori.
VALUE_MIN	L'entità fa riferimento al minimo valore numerico, tipicamente associato agli indicatori.
VOM_COST	L'entità fa riferimento all'indicatore <i>Variable Operation & Maintenance Cost</i> .
WAFER_THICKNESS	L'entità fa riferimento all'indicatore <i>Wafer Thickness</i> .

Relazione 1	Prima entità	Seconda entità
INDICATORE_DI_COAL Collega gli indicatori alla tecnologia del COAL o alle sue varianti	FOM_COST, ENERGY_INPUT_COST, INSTALLED_CAPACITY, LIFETIME, OM_COST, EMISSIONS, TEMPERATURE, FUEL_COST, DECOMMISSIONING_COST, PM_EMISSIONS, VOM_COST, ELECTRICITY_COST, CO2_EMISSIONS, MARKET_SHARE, SO2_EMISSIONS, EXTERNAL_COST, NUMBER_OF_UNITS, ENERGY_EFFICIENCY, GHG_EMISSIONS_COST, PRESSURE, CAPACITY_FACTOR, SYSTEM_COST, ELECTRICITY_SHARE, TYPICAL_SIZE, NOX_EMISSIONS	COAL_PC_AUSC, COAL_PC_SC, COAL_IGCC, COAL_PC, COAL_PC_SUB, COAL, COAL_PC_USC

Relazione 2	Prima entità	Seconda entità
INDICATORE_DI_CSP Collega gli indicatori alla tecnologia del CSP o alle sue	HEAT_STORAGE_MEDIA, ELECTRICITY_SHARE, OM_COST, SOLAR_MULTIPLE, ELECTRICITY_COST, CAPACITY, FINANCING_COST, MARKET_SHARE, CAPACITY_FACTOR, CONSTRUCTION_TIME, INSTALLED_CAPACITY, CYCLE_PRESSURE, HTF_TEMPERATURE, TOTAL_CUMULATIVE_CAPACITY,	CSP_VAR_LFR_STEAM_WS, CSP_VAR_CR_SALT_WS, CSP_VAR_CR, CSP_VAR_CR_STEAM_NS, CSP_VAR_PT, CSP, CSP_VAR_PT_OIL_WS, CSP_VAR_LFR,

varianti	STORAGE_CAPACITY, TEMPERATURE, LIFETIME, NOMINAL_EFFICIENCY, ANNUAL_CONVERSION_EFFICIENCY, ENERGY_EFFICIENCY, CYCLE_TEMPERATURE, CONCENTRATION_RATIO, CAPACITY_UNDER_CONSTRUCTION, SYSTEM_COST, LAND_USE, HTF_TYPE, STORAGE, TYPICAL_SIZE, CYCLE_TYPE, NUMBER_OF_UNITS, PLANNED_CAPACITY	CSP_VAR_LFR_STEAM_NS, CSP_VAR_PT_OIL_WWS, CSP_VAR_PT_SALT_WS, CSP_VAR_PT_STEAM_NS, CSP_VAR_DISH, CSP_VAR_PT_OIL_NS
----------	--	--

Relazione 3	Prima entità	Seconda entità
INDICATORE_DI_PV collega gli indicatori alla tecnologia del PV o alle sue varianti	MARKET_SHARE_PROJECTION, WAFER_THICKNESS, CELL_EFFICIENCY, ANNUAL_CONVERSION_EFFICIENCY, MODULE_EFFICIENCY, AVOIDED_CO2_EMISSIONS, COMMERCIAL_MODULE_EFFICIENCY, CAPACITY, EMISSIONS, MARKET_SHARE, BOS_COST, NOMINAL_EFFICIENCY, INSTALLED_CAPACITY, COST_BREAKDOWN, CO2_EMISSIONS, OM_COST, MATERIAL_USE, CAPACITY_FACTOR, MODULE_COST, LIFETIME, ENERGY_PAYBACK_TIME, ELECTRICITY_COST, SYSTEM_COST, AVAILABILITY_FACTOR, TYPICAL_SIZE, ENERGY_EFFICIENCY, ELECTRICITY_SHARE, TOTAL_CUMULATIVE_CAPACITY, LAND_USE	PV_VAR_NOVEL_OSC_DSSC, PV_VAR_NOVEL_OSC_OPV, PV_VAR_NOVEL_CPV, PV_VAR_TF_CIGS, PV_VAR_NOVEL_OSC, PV_VAR_NOVEL_AITF, PV_VAR_TF_AS, PV_VAR_CSI, PV_VAR_NOVEL_PV_VAR_TF_MJS, PV, PV_VAR_TF, PV_VAR_CSI_MC, PV_VAR_TF_CT, PV_VAR_CSI_SC

Relazione 4	Prima entità	Seconda entità
UNITA_DI collega l'unità di misura al relativo valore	MEASUREMENT_UNIT	VALUE, VALUE_MIN, VALUE_MAX

Relazione 5	Prima entità	Seconda entità
VALORE_DI collega il valore al relativo indicatore	VALUE, VALUE_MAX, VALUE_MIN	MARKET_SHARE, HTF_TEMPERATURE, AVAILABILITY_FACTOR, LAND_USE, CAPACITY_UNDER_CONSTRUCTION, CAPACITY, ELECTRICITY_SHARE, PLANNED_CAPACITY, ENERGY_PAYBACK_TIME, MODULE_EFFICIENCY, NOMINAL_EFFICIENCY, OM_COST, FINANCING_COST, LIFETIME, ENERGY_INPUT_COST, BOS_COST, INSTALLED_CAPACITY, WAFER_THICKNESS, SO2_EMISSIONS, MARKET_SHARE_PROJECTION, VOM_COST, TYPICAL_SIZE, EMISSIONS, PM_EMISSIONS, TOTAL_CUMULATIVE_CAPACITY, ANNUAL_CONVERSION_EFFICIENCY, STORAGE_CAPACITY, TEMPERATURE, DECOMMISSIONING_COST, ELECTRICITY_COST, GHG_EMISSIONS_COST, COMMERCIAL_MODULE_EFFICIENCY, SYSTEM_COST, HEAT_STORAGE_MEDIA, EXTERNAL_COST, CONCENTRATION_RATIO, CONSTRUCTION_TIME, FOM_COST, AVOIDED_CO2_EMISSIONS, NOX_EMISSIONS, CYCLE_TEMPERATURE, MODULE_COST, FUEL_COST, STORAGE, CYCLE_TYPE, SOLAR_MULTIPLE, MATERIAL_USE, COST_BREAKDOWN, CELL_EFFICIENCY, CO2_EMISSIONS, CYCLE_PRESSURE, HTF_TYPE, PRESSURE, NUMBER_OF_UNITS, UNITS_IN_OPERATION, ENERGY_EFFICIENCY, CAPACITY_FACTOR

Pre-annotazione e annotazione dei documenti di training e training del modello

Una volta definiti e aggiunti gli asset, si è proceduto alla pre-annotazione dei documenti di training. Nella Fase 1 la pre-annotazione è stata effettuata partendo dai dizionari e da un modello Rule Based appositamente creato per definire le entità VALUE, VALUE_MAX e VALUE_MIN, necessarie per categorizzare rispettivamente i valori tipici, massimi e minimi delle entità indicatori (i parametri tecnici ed economici che caratterizzano la tecnologia). Terminata la pre-annotazione, per ogni set di documenti di training è stato creato il relativo Annotation Set, una copia dei set pre-annotati destinati ad essere lavorati dagli annotatori umani. In Fase 1 del progetto l'annotazione degli otto Annotation Set di training è stata suddivisa fra due annotatori identificati nelle persone di due Senior Expert IBM. L'annotazione da parte dell'annotatore umano prevede l'eventuale correzione e/o integrazione delle entità e relazioni annotate automaticamente da Watson Knowledge Studio tramite la pre-annotazione, oltre all'eventuale annotazione della co-reference che permette al modello di capire se le entità all'interno del documento si riferiscono allo stesso concetto. Nella Figura 8 sono osservabili una serie di entità estratte da un documento di training

della Fase 1. Dopo aver completato e confermato l'annotazione degli Annotation Set di training, è stato eseguito il primo training del modello sugli otto Annotation Set di training.

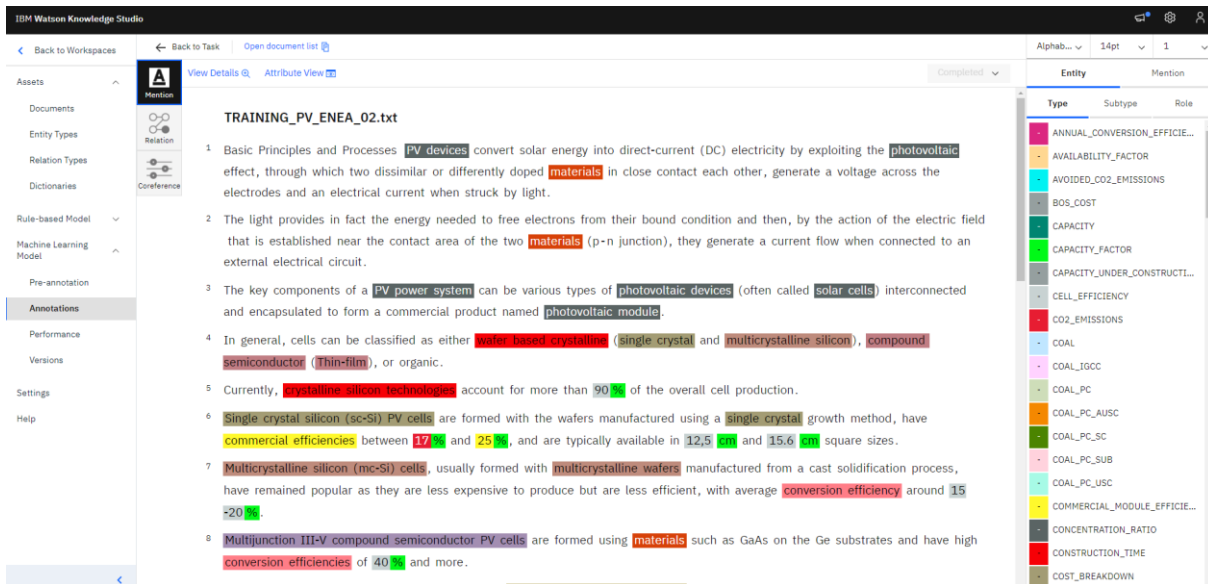


Figura 8 - Annotazione di un documento di training

Valutazione del modello

Terminato il training del modello è stata eseguita una valutazione delle sue prestazioni tramite degli appositi test. In questa fase sono stati utilizzati alcuni dei documenti (documenti di origine) forniti dal Committente nella fase iniziale, non utilizzati per l'addestramento e appositamente accantonati per essere utilizzati nella fase di valutazione (tre **set di test**). La valutazione prevede in sostanza la ripetizione sui set di test della procedura appena descritta. In particolare:

- la pre-annotazione dei set di test con il modello Machine Learning appena creato;
- la creazione degli Annotation Set di test, associati ai set di test pre-annotati;
- l'assegnazione degli Annotation Set di test agli annotatori umani;
- l'annotazione degli Annotation Set di test da parte degli annotatori umani;
-
- la valutazione del modello su tali Annotation Set di test.

La Figura 9 mostra le performance del modello creato nella Fase 1 del progetto. In generale, ove i risultati ottenuti non fossero soddisfacenti, si può procedere ad un eventuale ulteriore training del modello per ottenere risultati migliori.

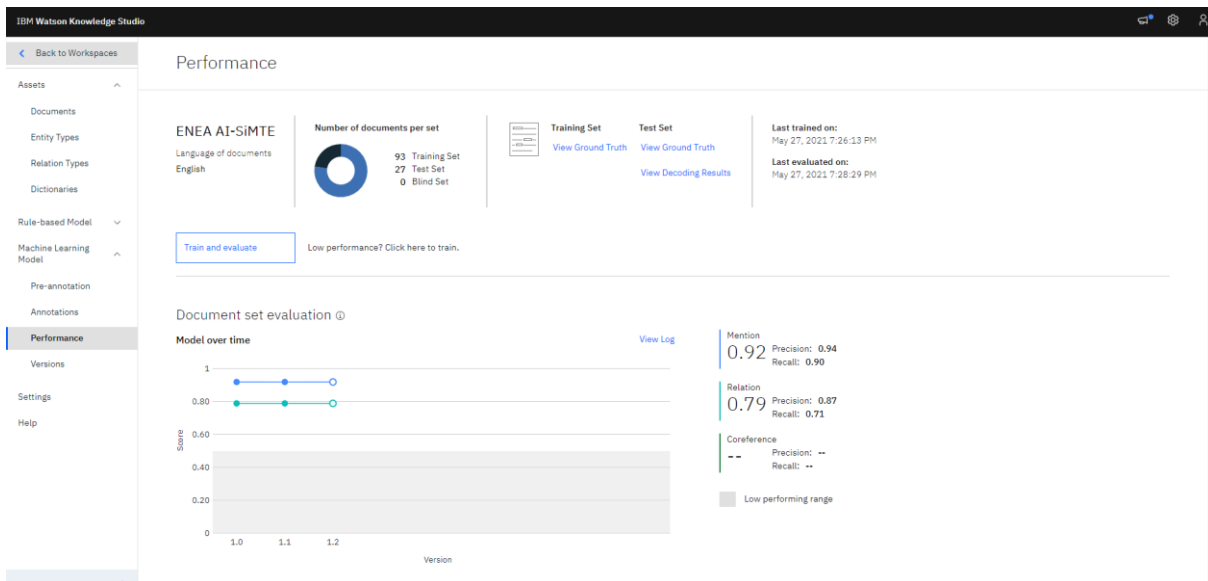


Figura 9 - Performance del modello cognitivo

Pubblicazione del modello su Watson Discovery

Nella Fase 1 del progetto, il modello generato dal componente Watson Knowledge Studio è stato “pubblicato” nel componente Watson Discovery, al fine di arricchire i documenti caricati in Watson Discovery con le entità e le relazioni sulle quali il modello è stato trainato. Il componente Watson Discovery è uno strumento di ricerca ed analisi basato sull’intelligenza artificiale, che estrae informazioni da ogni tipo di documento. In aggiunta a ciò, Watson Discovery conserva i documenti e i relativi arricchimenti come sentiment e concetti chiave, frutto delle analisi cognitive effettuate, e permette di effettuare query immediate in linguaggio naturale o linguaggio specifico Discovery, direttamente sui documenti. Il componente Watson Discovery è stato collegato anche al componente Watson Assistant e, tramite quest’ultimo, è stato utilizzato per permettere la lettura dei documenti in linguaggio naturale e in linguaggio Discovery.

Analisi e arricchimento cognitivo dei documenti con Watson Discovery

I tre documenti su Photovoltaics, Concentrating Solar Power e Coal, selezionati come documenti di riferimento per la Fase 1, come documenti costituenti la fonte da cui reperire le informazioni da fornire all’utente finale tramite l’assistente virtuale, sono stati caricati tramite l’interfaccia grafica del sistema in una *collection* di progetto, appositamente creata in lingua inglese. Watson Discovery ha quindi effettuato, in autonomia, analisi cognitive sui documenti, e li ha arricchiti con degli insight di default, fra cui, ad esempio, *keyword*, *sentiment*, e *category classification*. Alla collection di progetto è poi stato associato il modello Machine Learning generato da Watson Knowledge Studio, al fine di arricchire i documenti con due ulteriori tipologie di insight: le entità e le relazioni. Terminata l’associazione del modello, Watson Discovery ha infatti effettuato un’ulteriore analisi cognitiva per estrarre dai testi dei documenti le entità e le relazioni definite dal suddetto modello Machine Learning.

Training del SW cognitivo sulla struttura dei documenti e acquisizione delle tabelle dati

Una volta concluse le varie fasi fin qui descritte, è stato effettuato un apposito training del SW cognitivo con l’obiettivo di far comprendere a Watson Discovery la struttura dei documenti SiMTE. I documenti

disponibili sulla piattaforma SIMTE che caratterizzano le varie tecnologie energetiche hanno infatti un format specifico costituito da:

- un Abstract (Highlights) con sezioni caratterizzate da titoli standard (Process, Technologies and Market Aspects, Performance and Costs, Potential and Barriers);
- un testo con figure e tabelle diviso in capitoli e paragrafi, che in molti casi riproducono in forma più estesa la struttura dell'Abstract;
- una tabella finale che riassume per la tecnologia in esame e per le sue varianti i dati di maggior rilievo riguardanti prestazioni, costi e aspetti di mercato;
- una bibliografia.

Tale struttura dei documenti, pensata non specificatamente in vista dell'addestramento del SW cognitivo, ha favorito tuttavia l'addestramento in misura significativa, consentendo di ottenere risultati soddisfacenti anche nei ridotti tempi a disposizione per l'esecuzione delle attività. In particolare, come evidenziato in Figura 10, Watson Discovery è stato addestrato con l'obiettivo di identificare, all'interno dei documenti, parti dei documenti stessi, quali il titolo, l'autore, la data in cui il documento è stato redatto, i titoli dei capitoli, quelli dei paragrafi, il footer, le immagini, le tabelle, le note a piè di pagina, e le caption di tabelle e immagini. La corretta definizione della struttura dei documenti ha permesso la successiva suddivisione degli stessi per sottotitoli.

La struttura formattata dei documenti SIMTE è stata di aiuto anche nella fase di affinamento della risposta del sistema a interrogazioni (query) in linguaggio naturale. È stato infatti possibile istruire il sistema ad associare alla specifica query solo gli estratti dei documenti rilevanti e più in linea con la query stessa (come osservabile nella Figura 11).

Per quanto riguarda infine l'estrazione delle tabelle riassuntive finali dei documenti SIMTE (e dei dati in esse contenuti che vanno a costituire parte della banca dati SIMTE), trattandosi di tabelle non strutturate che possono contenere sia dati numerici che brevi testi e note di commento e precisazione, è stato necessario ricorrere ad una funzionalità Beta (non ancora commercializzata) di Watson Discovery. Tale funzionalità sfrutta un modello di Machine Learning pre-trainato incluso in Watson Discovery per trasformare le tabelle presenti nei documenti in file in formato JSON, che vengono poi lavorati dall'orchestratore. Questo meccanismo consente l'acquisizione delle tabelle e il trasferimento dei dati alla banca dati SIMTE previa verifica di correttezza operata da un supervisore umano.

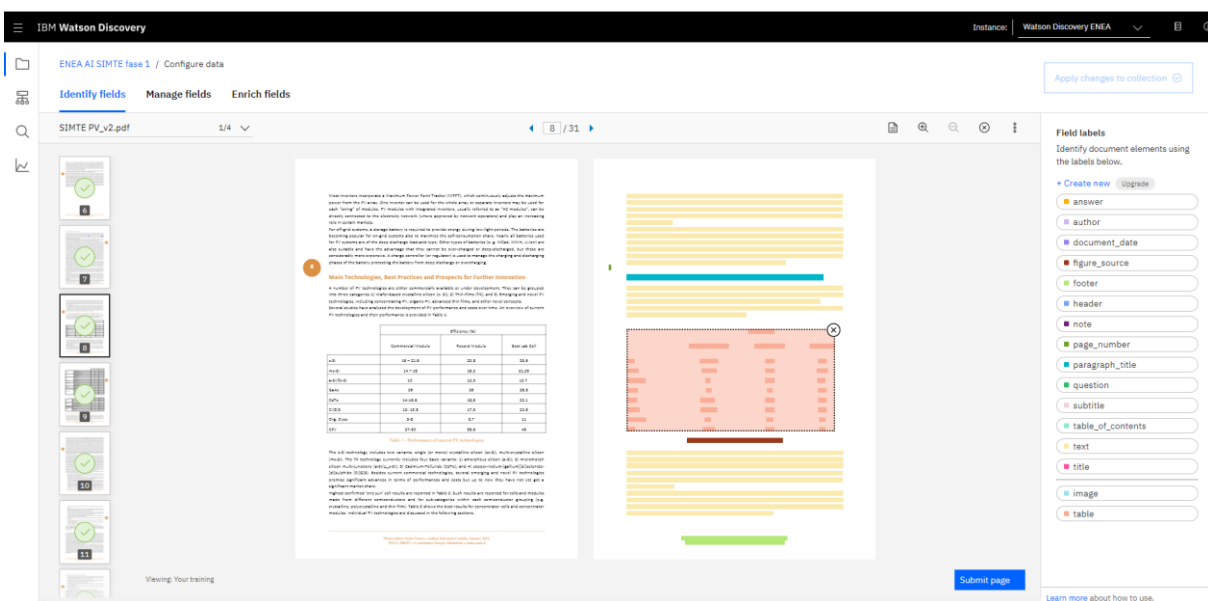


Figura 10 - Training strutturale dei documenti.

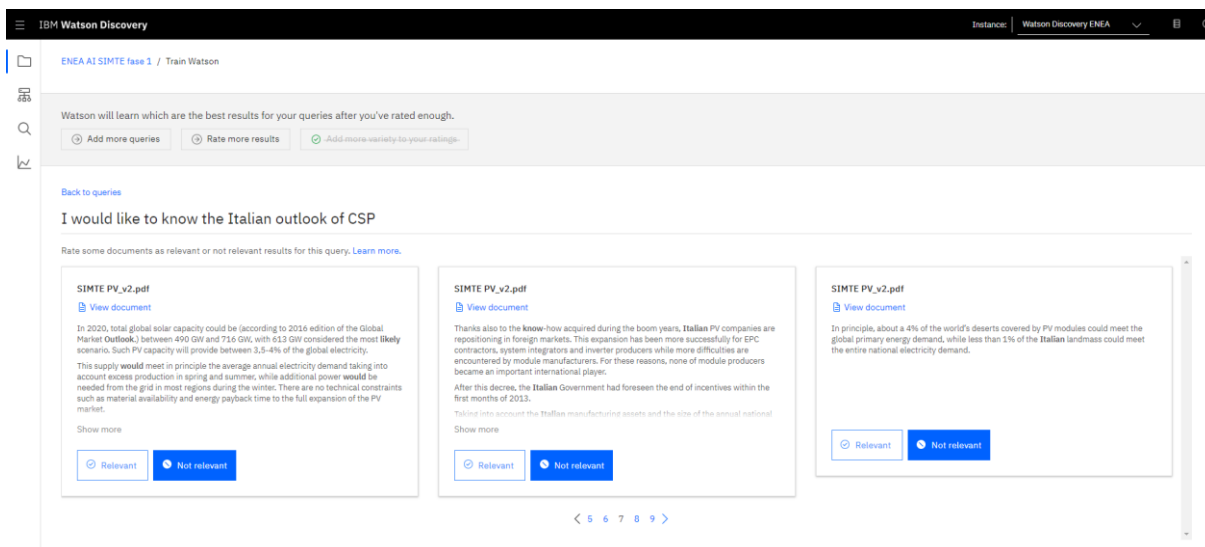


Figura 11 - Training query in linguaggio naturale.

2.3 Risultati della Fase 1

Al termine della Fase 1, il Fornitore Sistemi Informativi/IBM ha fornito al Committente ENEA l'accesso ai servizi sviluppati su IBM Cloud e ai vari componenti che costituiscono il sistema e una dimostrazione delle varie funzionalità del SW cognitivo relativamente ai tre cluster tecnologici oggetto di addestramento.

I test di funzionalità sono stati eseguiti congiuntamente dal Fornitore e dal Committente in data 17/06/2021 secondo i criteri stabiliti in fase di raccolta requisiti e previsti per la fine della Fase 1 nelle Specifiche Tecniche del contratto stipulato (Allegato 3, pagg. 2 e 4). I test hanno riguardato in particolare:

- la capacità di acquisizione dei documenti (lettura automatica ed estrazione dei dati di interesse, con trasferimento alla banca dati SiMTE, relativamente alle tecnologie oggetto di addestramento;
- la capacità dell'assistente virtuale di rispondere in modo congruo (con congruenza almeno nel 60% dei casi) a domande poste dall'utente in linguaggio naturale sulle tecnologie oggetto di addestramento;
- la capacità di fornire, a richiesta dell'utente, brevi report con testo, dati e tabelle estratti dai documenti presenti sulla piattaforma, in base alle domande formulate dall'utente stesso.

Infine, in data 22/06/2021, è stato svolto un training per il personale ENEA dedicato alla presentazione in dettaglio delle componenti cognitive della soluzione, Watson Assistant, Watson Discovery e Watson Knowledge Studio e ad una prima introduzione alle tecniche di addestramento.

I test effettuati in data 17/06/2021 hanno fornito esiti soddisfacenti e incoraggianti ai fini del passaggio alla prevista Fase 2 del contratto, dimostrando la potenzialità del sistema (ancorché creato in soli due mesi) di rispondere alle esigenze della piattaforma SiMTE, pur nella necessità di una ulteriore messa a punto. È risultato altresì evidente che i tempi disponibili e le risorse umane ed economiche allocate per l'attività contrattuale (sia dal Committente che dal Fornitore) non avrebbero permesso di applicare il sistema a tutte le circa 100 tecnologie energetiche contenute nella piattaforma SiMTE (obiettivo peraltro non previsto nelle Specifiche Tecniche). È stato pertanto deciso in base al piano di lavoro si passare alla Fase 2.

Allegati

Allegato 1: RELAZIONE TECNICA - Appalto per fornitura di servizi di implementazione/addestramento e diritti di accesso (license & cloud) a SW-AI (SoftWare – Artificial Intelligence) per applicazione a piattaforma SiMTE-ENEA –RTI Sistemi Informativi/IBM e VAR Group -

Allegato 2 : RAPPORTO FINALE DI ATTIVITA per servizi di implementazione/addestramento e diritti di accesso (license & cloud) a SW-AI (SoftWare – Artificial Intelligence) per applicazione a piattaforma SiMTE-ENEA - Sistemi Informativi/IBM

Allegato 3: Specifiche Tecniche per l'acquisto di servizi di implementazione/addestramento e diritti di accesso (license & cloud) a SW-AI (Software - Artificial Intelligence) per applicazione alla piattaforma SiMTE (Sistema informativo e di Monitoraggio delle Tecnologie Energetiche), ENEA – Agosto 2020