

REPORT DI PROGETTO

“*LABORATORIO SENTENZE*”: un esperimento di intelligenza artificiale

REFERENTI DI PROGETTO	2
Origine del progetto nella cornice di <i>Next Generation UPP</i>	3
Avvio della sperimentazione	4
Obiettivi della sperimentazione	6
Metodologia di lavoro	7
Milestone della sperimentazione	7
Test esplorativi	8
Primi risultati operativi e prospettive	12
Questioni aperte	15

REFERENTI DI PROGETTO

UniTo:

- Prof. Roberto Cavallo Perin
- Dott.ssa Eugenia Jona
- Dott.ssa Emma Scapino

CSI Piemonte:

- Data Strategy, Analytics, Geo Services
 - Aldo Passarelli - Coordinator
 - Angela Appendino - Project Manager
 - Francesca Ansaldi – Business Analyst
 - Michaela Jackson - Data Specialist
 - Maurizio Taddeo - Solution Designer
 - Alexander Fazari - Consultant
- A.I. e Robotic Process Automation
 - Claudio Parodi - Coordinator

Origine del progetto nella cornice di *Next Generation UPP*

La sperimentazione avviata dal Team del Prof. Roberto Cavallo Perin, docente ordinario di diritto amministrativo presso il Dipartimento di Giurisprudenza di UniTo, in collaborazione con il CSI Piemonte, nasce all'interno del più ampio Progetto "*Next Generation UPP: nuovi schemi collaborativi tra Università e uffici giudiziari Per il miglioramento dell'efficienza e delle Prestazioni della giustizia nell'Italia del nord-ovest*".

Com'è noto, *Next Generation UPP* ambisce a contribuire alla riduzione della durata media dei procedimenti e all'abbattimento dell'arretrato giudiziario. Il progetto intende infatti migliorare le prestazioni della giustizia dell'Italia del nord-ovest grazie al potenziamento degli Uffici per il Processo, la sperimentazione di nuovi schemi collaborativi tra le Università e gli Uffici giudiziari, e l'innovazione tecnologica (<https://www.nextgenerationupp.unito.it/home>).

In tale contesto, il Team del Prof. Cavallo Perin ha intrapreso una collaborazione con il CSI Piemonte, tenuto conto altresì di pregresse esperienze lavorative su progetti inerenti alla realizzazione di servizi di analisi descrittive e predittive, mediante l'utilizzo di tecniche di intelligenza artificiale, per quanto concerne le carriere degli studenti universitari, e alla collaborazione con l'Agenzia Piemonte Lavoro riguardo all'occupabilità dei soggetti che affluiscono ai Centri per l'impiego.

Il contributo domandato al CSI Piemonte è finalizzato a determinare la fattibilità di specifici tasks e a definire una metodologia di lavoro, nonché la realizzazione dei suddetti tasks, con messa in rilievo dei primi risultati operativi della sperimentazione in oggetto.

Tale sperimentazione, poi denominata Progetto "Laboratorio Sentenze", ha preso le mosse dalla condivisione con il CSI Piemonte di sentenze in materia di servizi pubblici, già riassunte e schematizzate dal Team del Prof. Cavallo Perin.

Ad oggi, si utilizza una porzione di sentenze del Consiglio di Stato, sempre in tema di servizi pubblici, accessibili sul sito Giustizia Amministrativa.

La sperimentazione mira allo sviluppo di un algoritmo di analisi testuale delle sentenze da inserire in un machine learning, che, in base al data lake composto da sentenze, individui l'esito di precedenti giudizi relativi a casi simili a quello in ordine al quale il giudice è chiamato a decidere.

Specificamente, il Progetto "Laboratorio Sentenze" si propone lo sviluppo di un algoritmo capace di generare sintesi di sentenze e di raffrontarle per poter, poi, riscontrare la similarità dei fatti alle stesse sottesi.

Ulteriore scopo del progetto consiste nel drafting assistito di sentenze, ossia nella creazione di modelli-template di sentenze per casi ripetitivi.

Ci si propone anche di esplorare la fattibilità della predizione dell'esito delle controversie, tenendo conto di tutte le implicazioni che questo approccio può comportare.

Lo scopo del Laboratorio consiste nella definizione di una metodologia da applicare, poi, alla progettazione di servizi di giustizia predittiva. In quest'ambito, il concetto di giustizia predittiva può intendersi come impiego di strumenti tecnologici evoluti sì da fornire un supporto al giudice nella fase decisionale e ottimizzare altresì i tempi della giustizia amministrativa.

La complessità delle decisioni giudiziarie e la necessità di analizzare un vasto volume di dati motivano l'adozione di tecnologie avanzate. I benefici attesi comprendono l'ottimizzazione dei tempi di ricerca e verifica delle sentenze, nonché il supporto nel processo decisionale. La metodologia è volta a supportare la sostenibilità e la scalabilità del sistema, assicurando al contempo la sicurezza e la privacy dei dati giudiziari.

Il laboratorio mira a definire una metodologia per la progettazione di servizi di giustizia predittiva, sfruttando l'intelligenza artificiale. In tale contesto, consideriamo il concetto di giustizia predittiva come impiego di questi strumenti tecnologici evoluti per fornire supporto alle decisioni giudiziarie ed ottimizzare i processi amministrativi.

Avvio della sperimentazione

La sperimentazione ha preso le mosse da un insieme di sentenze in materia di servizi pubblici condivise dal Team del Prof. Cavallo Perin con il Team di CSI Piemonte. L'attività su tali sentenze è stata svolta dalla dott.ssa Camilla Jacod, nell'ambito del Progetto *“Servizi pubblici e intelligenza artificiale –Anno 2021”*¹.

- Specificamente, le sentenze concernono le seguenti categorie di servizi pubblici: servizi pubblici locali di interesse economico generale
- compensazioni e obblighi di servizio pubblico
- concessione di servizi pubblici
- modalità di gestione dei servizi pubblici (affidamento *in house*)

¹ Il responsabile scientifico della borsa era il Prof. R. Cavallo Perin. La borsa è stata svolta dal 1 luglio 2021 al 30 settembre 2021 dalla dr.ssa Camilla Jacod e dal 30 settembre 2021 a 1 dicembre 2021 dalla dr.ssa Veronica Filippi.

- controllo analogo congiunto
- partecipazione di capitale privato nella società *in house*
- società miste

Lo schema delle sentenze - assunto come prototipo per lo sviluppo dell'algoritmo - si presenta nel seguente modo:

Estremi	Fatto	Decisum	Ratio Decidendi
Cons. St., sez. IV, 22 marzo 2021, n. 2426	Fiumicino Harbour Services Srl [appellante] contro il silenzio dell'Autorità Sistema Portuale del Mar Tirreno Centro Settentrionale a fronte dell'istanza della prima (concessionaria del servizio di raccolta e avvio al trattamento dei rifiuti delle navi scalanti nel Porto di Fiumicino) tesa a ottenere la riduzione tariffaria del servizio, con conseguente richiesta di risarcimento del danno per la riduzione della quantità di navi scalanti.	Rigetta l'appello e conferm a TAR Lazio-Roma, che aveva respinto il ricorso.	Il silenzio mantenuto dall'Autorità sulla domanda di revisione della tariffa è legittimo in quanto, il rischio operativo, nei rapporti di concessione di pubblico servizio, che ricomprende anche la fluttuazione della domanda, grava – a differenza che nell'appalto di servizi – strutturalmente sul concessionario, di talché la fluttuazione della domanda del servizio è insito nel meccanismo causale della concessione, che ha rapporti negoziali diretti con l'utenza, dalla cui richiesta trae la propria remunerazione. Conseguentemente, per farsi luogo a una revisione dei profili economici concordati, è necessaria la comprovata ricorrenza di eventi eccezionali e straordinari, oggettivamente esterni ed estranei al funzionamento del mercato di settore, non essendo invece sufficienti all'uopo mere fluttuazioni della domanda, dato fisiologico di ogni mercato, che l'operatore economico non può non considerare come aspetto caratterizzante, intrinseco ed ineliminabile del contesto in cui opera.

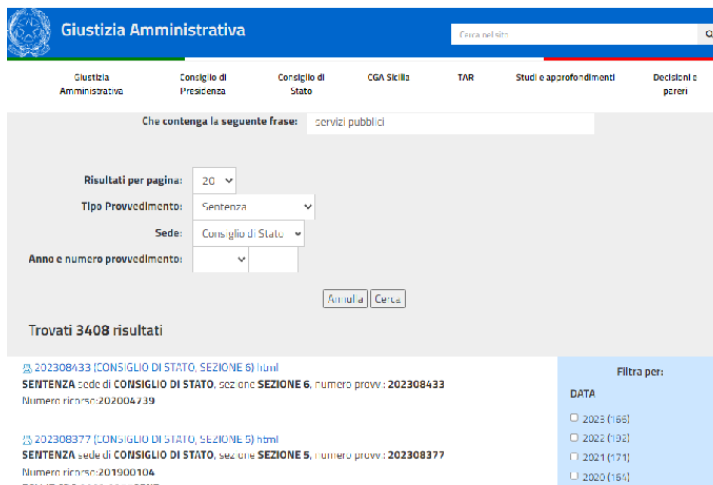
Tale schema di sentenza potrebbe rivelarsi utile per il lavoro di magistrati e avvocati:

- se generato in modo massivo (anche in forma più semplice) ai fini di un confronto agevole tra casi simili;
- se generato riguardo a determinate sentenze al fine di elaborarne una sintesi di veloce lettura.

Si evidenzia, tuttavia, come, nell'attuale fase della sperimentazione, il Team di CSI Piemonte si stia avvalendo delle sentenze del Consiglio di Stato, sempre in tema di servizi pubblici, accessibili al sito <https://www.giustizia-amministrativa.it/web/guest/dcsnpr.r>.

La scelta di procedere ad uno scarico massimo di giurisprudenza direttamente dal sito Giustizia Amministrativa è stata dettata dalla necessità di poter così disporre di un maggior numero di sentenze a disposizione.

Sperimentazione: primi test



The screenshot shows the 'Giustizia Amministrativa' website. At the top, there is a search bar with the text 'Cerca nel sito'. Below the search bar, there are several tabs: 'Giustizia Amministrativa', 'Consiglio di Piacenza', 'Consiglio di Stato', 'CGA Skills', 'TAR', 'Studi e approfondimenti', and 'Decisori e pcreri'. The 'Consiglio di Stato' tab is selected. Below the tabs, there is a search filter section. It includes a text input field with the value 'servizi pubblici'. Below this, there are several dropdown menus: 'Risultati per pagina:' (set to 20), 'Tipo Provvedimento:' (set to 'Sentenza'), 'Sede:' (set to 'Consiglio di Stato'), and 'Anno e numero provvedimento:' (set to a dropdown). There are also buttons for 'Annulla' and 'Cerca'. Below the search filter, it says 'Trovati 3408 risultati'. On the right side, there is a 'Filtra per:' section with a 'DATA' filter. It shows a list of years with the number of results in parentheses: 2023 (166), 2022 (192), 2021 (171), and 2020 (164). Below the search results, there are two sample results. The first is 'SENTENZA 202308433 (CONSIGLIO DI STATO, SEZIONE 6) html' with the text 'SENTENZA sede di CONSIGLIO DI STATO, sezione SEZIONE 6, numero provvi: 202308433' and 'Numero ricorso: 202004239'. The second is 'SENTENZA 202308377 (CONSIGLIO DI STATO, SEZIONE 5) html' with the text 'SENTENZA sede di CONSIGLIO DI STATO, sezione SEZIONE 5, numero provvi: 202308377' and 'Numero ricorso: 201900104'.

Elementi

Dati: ~1000 sentenze del Consiglio di Stato, contenenti la frase "**servizi pubblici**", dal 2016 al 2023.

Task: **predizione** dell'esito.

Strumenti: modelli del linguaggio di **vecchia generazione**.

Motivazioni: per **sondare** quanto effettivamente arduo sia il task, per **prendere confidenza** con i dati e per **stabilire una metodologia generale**.

Obiettivi della sperimentazione

1) Elaborazione di sintesi di sentenze:

- sintesi del fatto;
- generazione automatica di schemi suddivisi per fatto, diritto ed esito.

2) Riscontro della similarità tra fatti sottesi a precedenti giudiziari:

- ricerca di fatti (o sintesi) sottesi a casi simili a quello su cui il giudice è chiamato a pronunciarsi, così che egli possa conoscere le decisioni assunte in relazione ad essi e servirsene a supporto della propria, anche con riguardo ai motivi di diritto.

3) Predizione dell'esito di nuovi giudizi:

- studio sull'effettiva realizzabilità riguardo un modello di predizione degli esiti delle controversie.

Laboratorio Sentenze | Task da realizzare

Obiettivi del laboratorio

*Definire la **fattibilità** di alcuni task specifici ed individuare una **metodologia** di lavoro e di sviluppo, con prime evidenze derivanti dalla sperimentazione stessa, anche al fine di avere **contezza di quanto necessario** a livello tecnico, operativo, normativo e di eventuali **vincoli**, per una eventuale **fase realizzativa**.*

Task generali e casi d'uso individuati

1. **Sintesi** di sentenze/fatti e schematizzazione di fatto/diritto/esito
 - Sintesi dei fatti
 - Generazione automatica di schemi contenenti sintesi di fatto, diritto ed esito di sentenze
2. **Similarità tra fatti**
 - Ricerca di fatti o loro sintesi per ottenere sentenze simili al proprio caso e verificarne esiti a supporto della decisione, eventualmente anche principi di diritto
3. **Previsione** di classificazione delle sentenze in base all'esito
 - Drafting di sentenze che a partire da un fatto generi diritto ed esito POTENZIALI, che possa essere una base per chi deve poi scrivere la sentenza

Metodologia di lavoro

La metodologia di seguito proposta assume un ruolo centrale nell'approccio innovativo al settore della giustizia predittiva.

Attraverso una serie di passaggi, quest'ultima fornisce, per l'appunto, un metodo per l'analisi, l'elaborazione e, infine, la validazione dei dati in ingresso e degli output, sfruttando l'ausilio dell'intelligenza artificiale. In particolare, tale metodologia contempla l'analisi delle sentenze, l'estrazione dei dati, la definizione di obiettivi di elaborazione del linguaggio naturale, nonché di tipo predittivo, oltre che una rigorosa fase di validazione, evidenziando così l'importanza di un approccio strutturato e affidabile.

Milestone della sperimentazione

La sperimentazione riguarda l'elaborazione automatica di testi e pertanto necessita dell'impiego delle più recenti tecniche in ambito di Natural Language Processing.

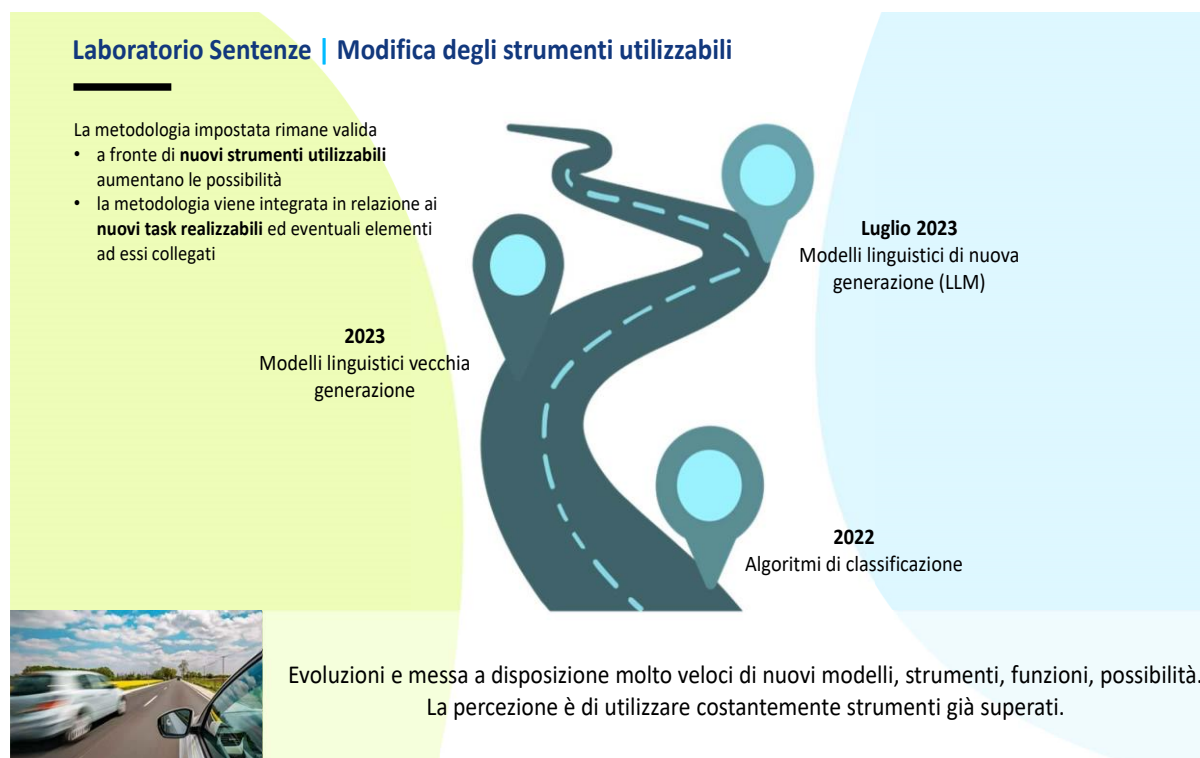
A partire dal 2017 circa, i modelli del linguaggio naturale più performanti si basano sull'architettura transformer: si tratta di un tipo di rete neurale basata sul "meccanismo dell'attenzione" in grado di mappare il linguaggio naturale su una forma matematica equivalente mantenendo il contesto.

I modelli del linguaggio naturale basati su tale tecnica si sono moltiplicati negli ultimi anni, ma è solo a partire dal 2022 che a livello internazionale si assiste ad un cambio di passo notevole

che vede l'avvento dei Large Language Model e delle Intelligenze Artificiali generative (agenti digitali) su di essi basate.

Tuttavia, ad inizio 2023 il contesto nazionale riguardo l'impiego dell'I.A. non permette ancora un utilizzo privo di "aree grigie". Pertanto, si è iniziato a sondare la fattibilità di alcuni task sfruttando modelli del linguaggio cosiddetti "di vecchia generazione", non generativi, ancorché basati sull'architettura transformer.

La sperimentazione iniziale si è basata su 1000 sentenze circa in materia di servizi pubblici. Inizialmente, al fine di esplorare, la possibilità di predizione dell'esito dei giudizi, si è fatto ricorso a modelli del linguaggio di precedenti generazioni, che, come si dirà in seguito, sono stati affiancati dall'ingresso di nuovi modelli del linguaggio, particolarmente utili per la comprensione/generazione di testo. Il primo step ha permesso di saggiare le complessità tecniche e di contesto, di acquisire familiarità coi dati e definire così una metodologia di lavoro generale.

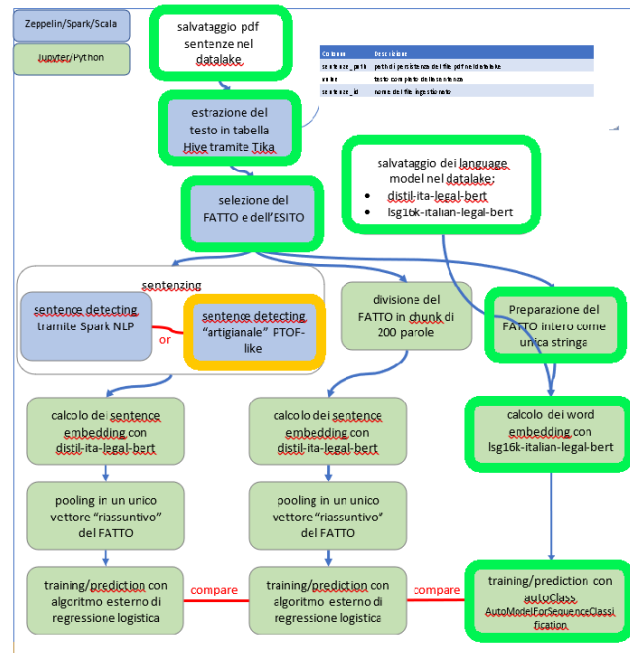


Test esplorativi

Nell'immagine sotto, una rappresentazione dello stato di avanzamento della sperimentazione a giugno 2023:

Dove siamo? Dove andiamo?

- Quali parti del **flusso** abbiamo **percorso**
- Documentazione** che traccia sperimentazione, metodologia e prospettive
- Principali **punti di attenzione**
 - Formato, Struttura** diversa delle **sentenze** provenienti da fonti diverse
 - capire fonti e struttura
 - Scarico di ulteriori sentenze pubbliche (anche tipologie diverse)
 - Elaborazione di sentenze immagini (senza parte laterale)
 - Sono stati elaborati **Fatto** insieme a **Diritto** (tagliando il testo alla lunghezza massima consentita dal modello)
 - si potrebbe capire se si riesce a trovare una logica per distinguerli
 - La **numerosità** delle sentenze e lo sbilanciamento tra i casi di ricorsi **ACCOLTI** e **RIFIUTATI** non sembrano ad ora sufficienti ad ottenere risultati accettabili
 - (rischio overfitting e non generalizzabili per ora)
- Prima prova di un **modello di sintesi**
 - si deve verificare tecnicamente il modello, allenarlo e si potrebbe usare per comparare i risultati con schemi sentenze e da verificare con esperti di dominio



Di seguito l'illustrazione delle fasi di lavoro:

1) FASE DI "INGESTION"

Le sentenze sono acquisite in diversi formati di documenti e vengono archiviate in un datalake messo a disposizione da CSI-Piemonte.

Proprietà	Dati	Diagramma ER	hive	stg_unito_unito_ml_fc_scienze_giuridiche	Tabelle	sentenze
Enter a SQL expression to filter results (use Ctrl+Space)						
Grid	sentenze_path	value	sentenze_id			
1	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	108/03/23, 10:18 N. 00569/2012 REG.RIC.11about:blank 1/4411N.	consstatosezv18dicembre2012n6488			
2	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:31 N. 04533/2012 REG.RIC.11about:blank 1/2911N.	consstatosezv2maggio2013n2390			
3	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04 about:blank11about:blank 1/2811Pubblicato il 3	consstatosezv31luglio2019n5444			
4	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:03 https://www.giustizia-amministrativa.it/portale/	consstatosezv15dicembre2020n8028			
5	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	107/03/23, 18:50 https://www.giustizia-amministrativa.it/portale/	consstatosezv21agosto2019n5778			
6	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:10 about:blank11about:blank 1/2611Pubblicato il 2	consstatosezv22ottobre2021n7093			
7	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	107/03/23, 11:04/23, 11:10 about:blank	io2017n2481			
8	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	108/03/23, 11:04/23, 11:10 about:blank	bre2014n4827			
9	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	e2014n5079			
10	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	re2017n4902			
11	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	108/03/23, 11:04/23, 11:10 about:blank	2015n3631			
12	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	aio2013n762			
13	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	io2015n2660			
14	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	2018n2583			
15	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	io2021n3682			
16	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	n1447			
17	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	bre2014n4599			
18	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	io2018n182			
19	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	re2016n5193			
20	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	io2018n655			
21	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	io2015n298			
22	hdfs://hdpha/datalake/UNITO/dataset_transformed/STAGE/unit	111/04/23, 11:04/23, 11:10 about:blank	io2013n1225			

2) FASE DI "PREPARATION"

Il contenuto delle sentenze viene estratto dai documenti tramite toolkit studiati per l'individuazione e l'estrazione di metadati e testo da molti formati differenti di documenti. In alcuni casi i metadati permettono di ottenere nell'immediato una divisione della sentenza nelle sue parti principali (FATTO; DIRITTO; P.Q.M.; ecc.); in altri occorre impostare un algoritmo di suddivisione.

Ai fini della ricostruzione dell'esito per la realizzazione del task di predizione, il Team di CSI ha proceduto all'analisi del P.Q.M., rilevando, però, come non sempre lo stesso abbia esito binario (0=respinto, 1=accolto), bensì presenti una difficoltà di individuazione per i non “addetti ai lavori”.

In taluni casi, il P.Q.M. include più volte i termini “*accolto*” e “*respinto*”: si è quindi prodotta una semplificazione matematica dei casi non binari, allo scopo di conteggiare il maggior numero di ambedue le casistiche.

Questa iniziale incertezza sulla determinazione dell'esito influisce fortemente sull'allenamento di modelli predittivi. L'avvento delle I.A. fornisce un supporto valido anche per questo genere di task.

Come anticipato sopra, le tecniche più recenti sull'elaborazione automatica dei testi prevedono la trasformazione del linguaggio naturale in una forma matematica equivalente, tramite un *embedding*.

In un primo tempo, sono stati testati alcuni language model di tipo BERT: un sentence transformer e una versione local-sparse-global16k di un modello BERT già allenato sul dominio legale. Questo nonostante fosse chiaro fin dal principio che il sentence transformer presentava un limite inderogabile di lunghezza del testo elaborabile e che l'LSG comportasse una diluizione eccessiva dell'informazione.

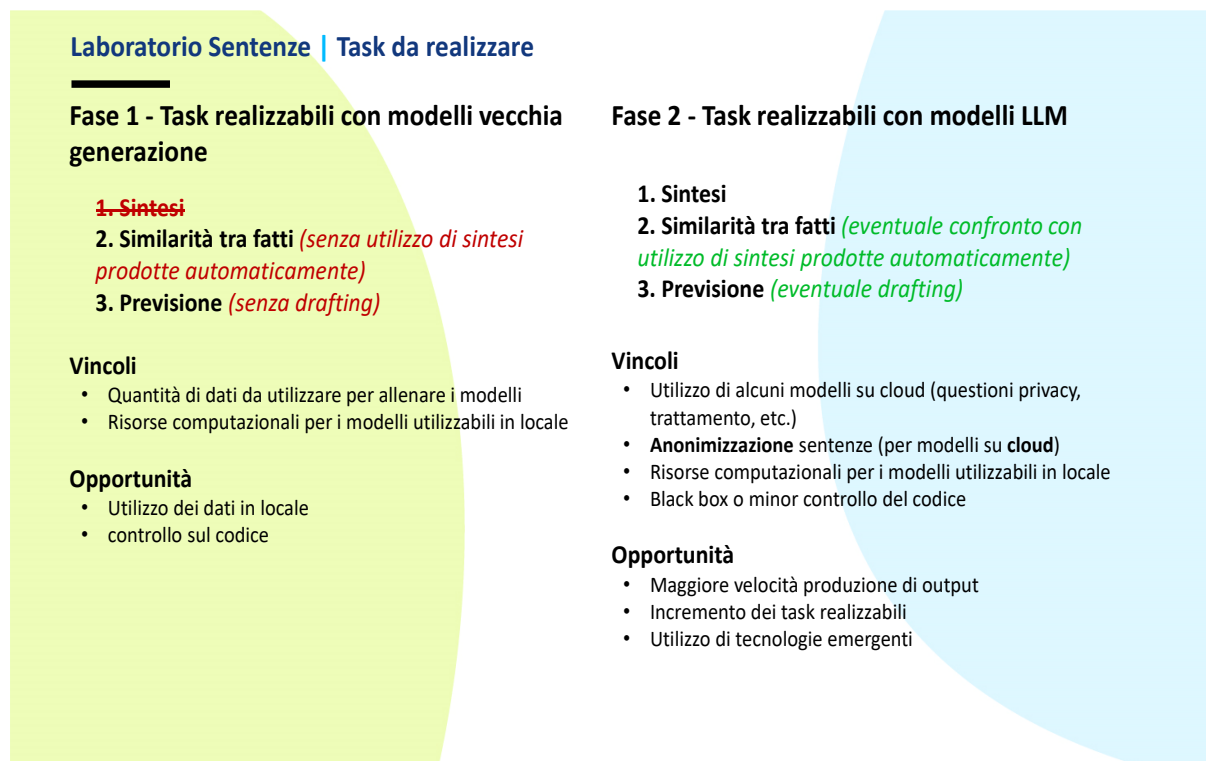
Ciononostante, è stato possibile condurre test ed è stata, così, impostata una metodologia di lavoro tutt'oggi valida.

Nel mese di luglio 2023, a fronte di nuovi strumenti utilizzabili, lo scenario si è ampliato: sono stati introdotti nuovi modelli generativi per la sintesi. In conseguenza, la metodologia di lavoro è stata integrata per adattarla ai nuovi tasks realizzabili. Ciò a riprova che il campo dell'intelligenza artificiale è un ambito in costante evoluzione.

Con riferimento alla sperimentazione in oggetto, l'avvento di modelli di linguaggio di ultima generazione ha determinato un notevole avanzamento in relazione sia al task di sintesi

automatizzata di fatti e sentenze stesse sia a quello di riscontro della similarità tra casi, nonché alla possibilità di individuare parti della sentenza, quali, ad esempio, il P.Q.M..

Nell'immagine sottostante si evidenzia che alcuni obiettivi di complessa o impossibile realizzazione con precedenti modelli, sono invece nativamente realizzabili con i modelli del linguaggio disponibili da luglio.



Si precisa che i modelli generativi del linguaggio costituiscono una forma di intelligenza artificiale addestrata per generare testo.

Il funzionamento dei medesimi consta di due fasi:

- 1) fase di domanda/prompt: l'utente formula una domanda/prompt da sottoporre al modello;
- 2) fase di risposta: il modello processa il prompt ed elabora una risposta sulla base di quanto ha appreso in fase di addestramento. La risposta prodotta dal modello si fonda sul contesto fornitogli in fase di domanda e su quanto dallo stesso appreso dai testi analizzati nel corso dell'addestramento.

Primi risultati operativi e prospettive

Nella fase iniziale della sperimentazione (quando, come scritto sopra, non si poteva accedere a modelli del linguaggio di ultima generazione), sono state effettuate delle prove in relazione al task di predizione dell'esito dei giudizi (*alias* prove di classificazione binaria dei ricorsi in "accolto" e "respinto"). In proposito, si evidenzia come siano state create diverse pipelines al fine di individuare la metodologia di lavoro più efficace per la realizzazione del suddetto task. Sono state sperimentate le seguenti pipelines:

- 1) Pipeline tf-idf su FATTO intero+classifier;
- 2) Pipeline chunk 200 sovrapposti+sentence transformer distil+classifier;
- 3) Pipeline frasi della class "popolo"+sentence transformer distil+classifier;
- 4) Pipeline testo intero con LSG16k+CLS+classifier;
- 5) Pipeline testo intero con LSG16k+AutoClass.

La pipeline n. 3 è quella che ha sortito i risultati migliori, benché non pienamente soddisfacenti. In particolare, le problematiche riscontrate attenevano alle insufficienti risorse computazionali per il caso del modello lsg16k, all'inadeguata quantità di dati a disposizione con riferimento a tutte le pipelines e all'incertezza nell'individuazione dell'esito per i dataset di training.

Sperimentazione: primi risultati

L'accuracy misura la capacità di avvicinarsi al risultato specifico. Può essere a scapito della precision: per esempio, si predicono correttamente tutti i respinti, ma a scapito di molti falsi respinti.

Prove di predizione (classificazione binaria dei ricorsi in accolti e respinti)				
ESPERIMENTO	DATI DI INPUT	ACCURACY	CONFUSION MATRIX	NOTE
NUMERO 1 - Pipeline tf-idf su FATTO intero + classifier	modello allenato su 1076 sentenze del Consiglio di Stato, contenenti la frase "servizi pubblici", dal 2016 al 2023	0,69	TP=44, FP=18, TN=39, FN=19	benchmark con tecnica tradizionale, nessun modello del linguaggio
NUMERO 2 - Pipeline chunk 200 sovrapposti + sentence transformer distil + classifier	modello allenato su 1076 sentenze del Consiglio di Stato, contenenti la frase "servizi pubblici", dal 2016 al 2023	0,64	TP=39, FP=23, TN=38, FN=20	quantità di dati sufficiente per questa pipeline? incertezza nell'individuazione dell'esito anche solo per il dataset di training
NUMERO 3 - Pipeline frasi della class "popolo" + sentence transformer distil + classifier	modello allenato su 1076 sentenze del Consiglio di Stato, contenenti la frase "servizi pubblici", dal 2016 al 2023	0,71	TP=44, FP=18, TN=42, FN=16	quantità di dati sufficiente per questa pipeline? incertezza nell'individuazione dell'esito anche solo per il dataset di training
NUMERO 4 - Pipeline testo intero con LSG16k + CLS + classifier	modello allenato su 1076 sentenze del Consiglio di Stato, contenenti la frase "servizi pubblici", dal 2016 al 2023	0,65	TP=39, FP=23, TN=40, FN=18	i dati sembrano essere troppo pochi per questa pipeline con modello LSG16k incertezza nell'individuazione dell'esito anche solo per il dataset di training
NUMERO 5 - Pipeline testo intero con LSG16k + AutoClass	modello allenato su poco più di 4000 sentenze del Consiglio di Stato, senza filtro sul contenuto	0,48		il dataset di training non è significativo i dati sembrano essere troppo pochi per questa pipeline con modello LSG16k incertezza nell'individuazione dell'esito anche solo per il dataset di training

Attraverso l'uso dei modelli del linguaggio di vecchia generazione si è, dunque, esaminata la possibilità di realizzazione del task relativo alla predizione dell'esito delle controversie giudiziarie.

Acquisito accesso ai modelli del linguaggio di nuova generazione, la sperimentazione si è, invece, focalizzata sulla fattibilità dei tasks di sintesi automatizzata del fatto (ed eventualmente dell'intera sentenza) e di riscontro automatizzato della similarità tra fatti.

Su detti task, data anche la velocità di nascita di nuovi modelli, sono in corso diverse prove, che mirano a confrontare i risultati di modelli specifici, utilizzabili sia in locale, sia su cloud e che permettano di identificare quelli più appropriati agli obiettivi di progetto, nel contesto di riferimento. L'utilizzo di modelli diversi permette anche di sperimentare una "validazione" incrociata dei risultati degli altri.

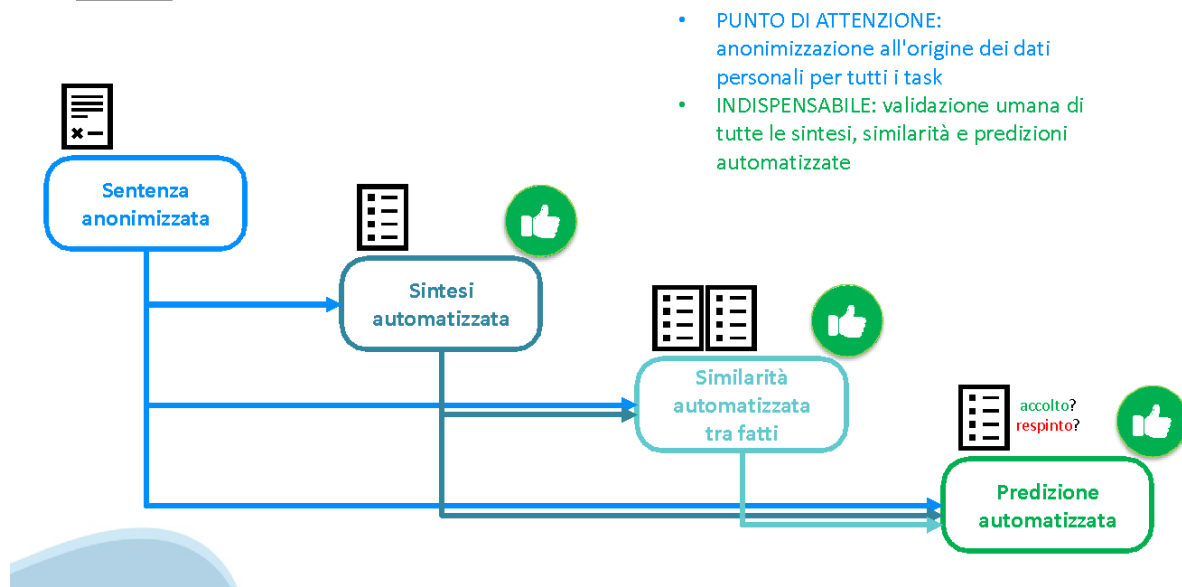
Ad oggi sono state svolte e sono in corso di affinamento/sviluppo le seguenti attività:

- anonimizzazione dei fatti, con differenziazione delle entità presenti nella stessa sentenza ai fini di non decontestualizzarne i contenuti, utili in particolare per i task di sintesi/similarità
- estrazione del fatto dal corpo della sentenza
- sintesi dei fatti con modelli in cloud e locali
- estrazione degli esiti dei ricorsi dal P.Q.M. (utile per i tasks di sintesi e predizione)
- estrazione di concetti/parole chiave dai fatti (utili per eventuali casi d'uso di ricerca, similarità dei fatti, pre-validazione delle sintesi)
- similarità dei fatti con indice di similarità
- predizione degli esiti con modelli "di prima generazione"

Attività da sperimentare:

- prove di validazione sintesi attraverso i concetti
- prove di validazione sintesi con cross model (usiamo un modello del linguaggio per validare le sintesi fatte con un altro modello)
- eventuale predizione degli esiti con modelli LLM (quelli disponibili da luglio)

Flusso attività



L'obiettivo più sfidante resta, dunque, quello della predizione automatizzata dell'esito dei giudizi. Questo obiettivo pone una serie di interrogativi noti riguardo l'effettiva praticabilità. E' possibile che il risultato finale, invece che portare alla classificazione dei casi in due categorie (accolti e respinti), si concretizzi nel suggerimento da parte di un modello predittivo su tre classi: le due classi degli accolti e dei respinti con ragionevole affidabilità (rispetto ad una confidenza da impostare a livello di modello), più la classe degli incerti, la zona grigia su cui il modello non riuscirà ad esprimersi in maniera che riteniamo sufficientemente attendibile e che sarà quella su cui il professionista dovrà concentrarsi.

Prima di raggiungere tale obiettivo, si impone un lavoro meticoloso sui tasks di sintesi e riscontro della similarità (sui quali, infatti, la sperimentazione si sta concentrando). Invero, per la predizione dell'esito del nuovo giudizio, l'algoritmo potrà attingere alle sentenze (o alle loro sintesi) emesse precedentemente su fatti simili. In conseguenza, tanto più gli algoritmi saranno accurati nell'elaborazione della sintesi e nel riscontro della similarità tra fatti quanto più i risultati inerenti al task di predizione automatizzata potranno essere soddisfacenti. Si precisa, tra l'altro, che le sintesi e i riscontri di similarità realizzati saranno validati da esperti della materia in modo avere il controllo sui risultati degli algoritmi.

Questioni aperte

Nell'ottica del conseguimento di risultati il più efficace possibile, rimangono, comunque, delle questioni sulle quali sarà opportuno un confronto tra i due teams, in particolare:

- i dati di cui è necessaria l'anonimizzazione;
- l'individuazione delle parti della sentenza rilevanti ai fini degli obiettivi condivisi (fatto; ratio decidendi; decisum; lessico giuridico ricorrente; norme citate);
- il numero di parole su cui si intende impostare la sintesi della sentenza (ad oggi, la sintesi della parte in fatto è impostata su 150 parole);
- la formulazione il più puntuale possibile delle domande richieste al modello. Quest'ultimo genera, infatti, un testo, in base alla domanda (c.d. prompt) formulata dall'utente. Il testo prodotto dipende fortemente dalle istruzioni fornite: è fondamentale formulare la domanda in modo puntuale e dettagliato, poiché il modello esegue pedissequamente la richiesta contenuta nel prompt. Occorre, dunque, evitare istruzioni ambigue e generiche; diversamente, la risposta potrebbe non essere quella ricercata;
- l'individuazione dei concetti/parole-chiave che è utile estrarre dal testo.