



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Corso di Laurea Magistrale in Data Science per l’Economia e le Imprese

Sviluppo di modelli per l’emotion recognition
di testi in lingua italiana

Development of models for emotion
recognition of Italian text

Relatore:

Prof. Potena Domenico

Tesi di Laurea di:

Ridolfi Lorenzo

Correlatore:

Prof. Mircoli Alex

Anno Accademico 2021 – 2022

Abstract

Il Natural Language Processing è una branca dell'intelligenza artificiale in grado di fornire ai computer l'abilità di elaborare il testo e le parole pronunciate in linguaggio naturale, rendendoli capaci di apprendere e ricavare significato da esso.

Nel contesto digitale attuale, la grande mole di dati a disposizione ha favorito lo sviluppo di applicazioni intelligenti in grado di compiere attività complesse che richiedono specifiche tecniche di elaborazione del linguaggio.

L'obiettivo del seguente lavoro consiste nel realizzare modelli di classificazione in grado di riconoscere le emozioni umane in contenuti testuali provenienti da social network e scritti in lingua italiana.

La fase sperimentale parte da un'analisi preliminare esplorativa e descrittiva, con l'intento di verificare la bontà del dataset e individuare eventuali pattern e relazioni nascoste nei dati. Segue poi l'implementazione di tecniche unsupervised come Word2Vec, un modello in grado di prendere una parola e restituire un vettore numerico associato alla parola stessa, così da apprendere le associazioni tra parole e identificare un livello di somiglianza semantica. Si conclude l'analisi mediante l'utilizzo di tecniche di clustering per raggruppare i word embedding ottenuti con Word2Vec.

L'analisi supervised sviluppa due classificatori delle emozioni tramite due differenti modelli pre-allenati, GPT-3 e Google BERT. Entrambi sono modelli di deep learning progettati per elaborare il linguaggio naturale ma differiscono sotto alcuni aspetti e per tale motivo sono messi a confronto.

Al termine della fase sperimentale, è possibile confermare che entrambi i classificatori raggiungono performance elevate nei dati di test, con livelli di accuracy pari a 90% con GPT-3 e 89% con BERT.

Sommario

1. INTRODUZIONE	6
1.1. CONTESTO DI RIFERIMENTO	6
1.2. MOTIVAZIONI	8
1.2.1. Sfide e opportunità	8
1.2.2. Struttura dell'elaborato	10
2. MODELLI TEORICI PER L'ANALISI DELLE EMOZIONI	12
2.1. GLI STUDI DI PAUL EKMAN	12
2.2. PLUTCHIK: LA RUOTA DELLE EMOZIONI	13
3. OBIETTIVI DI ANALISI	17
4. TECNOLOGIE UTILIZZATE	20
4.1. ARTIFICIAL INTELLIGENCE	20
4.1.1. Machine learning	22
4.1.2. Deep learning	23
4.1.3. Natural Language Processing	25
4.1.4. Transformers	26
4.2. UNSUPERVISED LEARNING	31
4.2.1. Word2Vec	31
4.2.2. K-means	35
4.3. SUPERVISED LEARNING	36
4.3.1. GPT-3	37
4.3.2. Google BERT	38
5. SET-UP SPERIMENTALE	42
5.1. IL DATASET	42
5.2. ANALISI DESCRITTIVA	44
5.3. UNSUPERVISED LEARNING	49
5.3.1. Word2Vec	49
5.3.2. Clustering	52
5.4. SUPERVISED LEARNING	56
5.4.1. GPT-3	56
5.4.2. Google BERT	59
5.5. VALUTAZIONE DEI RISULTATI OTTENUTI	61
6. CONCLUSIONI	65
6.1. Sviluppi futuri	66
SITOGRAFIA & BIBLIOGRAFIA	69

1. INTRODUZIONE

Nel periodo storico attuale, il proliferare di grandi quantità di dati in formato digitale ha permesso ad analisti ed esperti informatici di usufruire di tali risorse per produrre analisi accurate circa le attività principali che gli internauti svolgono in rete. Le suddette analisi hanno una molteplicità di scopi e risultano essere estremamente utili per acquisire informazioni di valore dal quale estrarne conoscenza.

Un ruolo fondamentale nel contesto appena citato viene assunto dai social media e in particolare dagli ormai noti social network, grandi piattaforme digitali che permettono la comunicazione e la condivisione di contenuti multimediali. Sono i servizi internet più diffusi dall'avvento del web 2.0 e sono stati creati sui principi ideologici e tecnologici di quest'ultimo.

1.1. CONTESTO DI RIFERIMENTO

Secondo i dati pubblicati dal Global Web Index, circa 4,6 miliardi di persone hanno un profilo social attivo nel 2022, con un incremento del 10% rispetto l'anno passato e con prospettive future ancora in crescita. In Italia invece, circa 43 milioni di persone sono attive sui social, più del 70% della popolazione e più dell'84% della popolazione di età pari o superiore a 13 anni [1].

Osservando i dati appena mostrati, è lecito supporre l'importanza che tali piattaforme assumono nell'universo del word wide web, rendendosi protagoniste assolute della rivoluzione digitale in atto ormai da anni.

Tra le piattaforme più acclamate non si può non citare Facebook con ben 2.9 miliardi di utenti, considerato il padre dei social network e al primo posto nella classifica mondiale per maggior numero di profili attivi. Seguono poi YouTube con 2.5 miliardi di utenti, l'applicazione di messaggistica istantanea WhatsApp che conta 2 miliardi di utilizzatori

e Instagram con 1.5 miliardi. Chiudono il club del miliardo WeChat e Tik-Tok, con rispettivamente 1.2 e 1 miliardo di utenti attivi nel 2022.

Ciò che contraddistingue un social network piuttosto che un altro sono le relazioni che si vengono a instaurare tra gli individui, gruppi e organizzazioni facenti parte del sistema, che uniti da interessi di varia natura creano community dove poter condividere contenuti multimediali inerenti al contesto di riferimento. La possibilità di creare veri e propri network di persone permette loro di sentirsi socialmente accettati e ciò genera in essi un senso di appartenenza tale da potersi esprimere al meglio, mostrando i propri sentimenti e le proprie emozioni senza alcun filtro sociale che spesso, anche inconsciamente, viene utilizzato nella vita reale con il fine di non apparire estranei all'ambiente circostante.

Ad oggi, sempre più persone sposano la filosofia dei social network e partecipano attivamente sfruttando le loro potenzialità. Tutto ciò genera una quantità di informazioni inimmaginabile: basti pensare che, secondo i risultati forniti da IBM, il 90% dei dati attualmente disponibili in rete sono stati generati solo negli ultimi 2 anni. Ciò è stato possibile grazie allo sviluppo di nuove tecnologie, internet, cloud, social media, social network e altri sistemi tecnologici all'avanguardia.

Nell'attuale era dei Big Data i numeri sono sorprendenti: 5,7 milioni di ricerche Google, 575.000 tweet pubblicati, 65.000 foto condivise su Instagram, 694.000 ore di streaming video su YouTube, 100.000 utenti connessi su Teams, 283.000\$ spesi su Amazon dai consumatori [2]. Tutto in un minuto di connessione internet.

La possibilità di accedere a tali informazioni ha favorito lo sviluppo di una serie di tecniche di analisi sofisticate rendendo più accurato il processo decisionale aziendale che, guidato dall'innovazione, crea esperienze migliori per clienti e dipendenti, abilita le persone a prendere decisioni di business più smart e permette loro di adattarsi al mutamento dell'ambiente circostante.

Sulla base di tale presupposto, il seguente progetto si è concentrato sull'analisi delle emozioni umane espresse nel testo mediante algoritmi di intelligenza artificiale.

1.2. MOTIVAZIONI

Il contesto sociale definito in precedenza, unito agli studi condotti in ambito emozionale, ha ispirato lo sviluppo del progetto in questione, che grazie alle conoscenze tecniche e alle tecnologie informatiche, ha analizzato nel dettaglio le emozioni umane in un'ottica digitale.

Come già riportato all'inizio del capitolo introduttivo, la digital transformation ha favorito lo sviluppo di tecnologie all'avanguardia in grado di apportare innovazione in differenti aree di business, e ciò ha richiesto l'implementazione di nuove strategie volte a raggiungere nuovi obiettivi.

1.2.1. Sfide e opportunità

L'innovazione digitale ha origine dai dati raccolti ed opportunamente organizzati ed analizzati, con il fine di tradurre le informazioni presenti in conoscenza di valore. Tale conoscenza permette alle imprese di sviluppare prodotti e servizi innovativi e coerenti alle richieste esigenti dei consumatori, sempre più selettivi nel loro processo decisionale.

Ascoltare il cliente è fondamentale in tutto il percorso di creazione del valore, ma essere in grado di cogliere particolari richieste non ancora rese esplicite mediante opportune analisi di contesto, permette di acquisire vantaggi competitivi durevoli e sostenibili nel tempo.

Se è vero che i dati sono considerati il “petrolio del ventunesimo secolo”, allora l'implementazione di tecniche di data analysis permettono di estrapolare da essi opportunità interessanti in svariati contesti economici. Per restare competitivi non è più possibile sottrarsi al processo di digitalizzazione, occorre quindi diventare data driven e dotarsi di sistemi capaci di raccogliere, organizzare ed analizzare dati.

Ad oggi è sempre più difficile immaginare scenari in cui le attività di data analysis non vengano applicate, e non solo in aziende grandi e strutturate, ma anche in quelle dalle dimensioni medio-piccole, che implementano le varie tecniche per diversi scopi [7].

Le opportunità derivanti da una corretta gestione del dato permettono di accedere ad informazioni preziose, in grado di rendere più consapevoli le organizzazioni nel momento in cui devono prendere decisioni importanti per il futuro delle stesse. L'accesso a internet rappresenta oggi il canale di comunicazione più diffuso al mondo, dove le imprese offrono beni e servizi altamente innovativi a un target per così dire globale, rispondendo efficacemente a richieste di personalizzazione, differenziando la propria offerta rispetto ai diretti concorrenti e migliorando di fatto la qualità del prodotto, reso grazie sfruttando le potenzialità dei dati digitali.

Ma i benefici derivanti dalle attività di data analysis e, più in generale, di data mining, non sono limitati al miglioramento dei prodotti o servizi per i clienti finali, ma investono le più svariate aree aziendali e, se adeguatamente realizzate, permettono cambiamenti sostanziali in grado di migliorare ogni aspetto del processo in esame. Nel dettaglio, i principali benefici derivanti da una corretta analisi possono consistere in miglioramenti in termini di efficienza, maggior controllo dei costi, risparmio di tempo, identificazione di trend progettuali ottimali, identificazione di problemi di performance, maggior qualità dei processi produttivi e delle soluzioni offerte.

La sfida, in quest'ottica, consiste nel saper estrarre valore dall'immensa quantità di dati presenti in rete mediante opportune metodologie di raccolta ed organizzazione dei dati. Gestire i dati raccolti, archivarli in modo corretto e sicuro, visualizzarli in maniera comprensibile e analizzarli mediante modelli predisposti così da rispondere a domande da parte del management, sono task complessi che per essere realizzati con successo devono essere sostenuti da un insieme di fattori chiave. Tra questi, dotarsi di infrastrutture adeguate a sostenere il ciclo di vita dei dati, di strumenti software all'avanguardia e persone con background tecnico in grado di manipolare i dati, garantiscono il raggiungimento dei risultati sperati.

1.2.2. Struttura dell'elaborato

Il presente elaborato è organizzato nei seguenti capitoli:

il secondo capitolo descrive i modelli teorici per l'analisi delle emozioni proposti da Paul Ekman e Robert Plutchik, fornendo una panoramica generale al problema di riconoscimento delle emozioni.

il terzo capitolo espone con chiarezza il contesto specifico di applicazione e gli obiettivi che si intendono raggiungere.

Il quarto capitolo propone una descrizione dettagliata circa le architetture di riferimento e le tecnologie utilizzate. Sarà descritto il dataset e spiegato il funzionamento dal punto di vista tecnico dei principali algoritmi utilizzati, nonché della metodologia adottata nelle diverse fasi progettuali. In particolare, verrà illustrata la logica con la quale le architetture sono state implementate.

Il quinto capitolo espone il set-up sperimentale, la valutazione degli esperimenti e la discussione tecnica dei risultati ottenuti. In particolare, verranno riportati gli esperimenti eseguiti in fase di test sia per la parte non supervisionata che per la parte supervisionata. Sarà esposta la valutazione sperimentale, il parameter tuning e le metriche utilizzate per definire i risultati finali. Un ulteriore paragrafo sarà dedicato invece alla discussione dei risultati ottenuti.

Il sesto ed ultimo capitolo riassume il lavoro svolto e ne evidenzia i tratti conclusivi. Alla luce dei risultati ottenuti, verrà illustrata la performance complessiva e verranno suggeriti spunti per alcuni sviluppi futuri interessanti.

2. MODELLI TEORICI PER L'ANALISI DELLE EMOZIONI

Esistono diverse teorie che spiegano il funzionamento delle emozioni negli esseri umani, ma attualmente quella più nota e dominante propone un sistema di classificazione di tipo categoriale in grado di distinguerle nettamente l'una dall'altra [3].

Sebbene questo tipo di approccio possa avere degli svantaggi in situazioni non proprio chiare in cui si potrebbero manifestare più emozioni contemporaneamente, diverse ricerche propongono una stratificazione omogenea delle stesse e permettono di studiarne le relazioni sottostanti.

In questa sede vengono discussi i contributi che Paul Ekman e Robert Plutchik hanno dato alla ricerca scientifica, proponendo strumenti analitici in grado di differenziare le diverse tipologie di emozioni e fornendo un quadro universale circa la loro manifestazione.

2.1. GLI STUDI DI PAUL EKMAN

Paul Ekman, psicologo statunitense noto per le sue ricerche scientifiche sulle espressioni facciali, è considerato un pioniere nel campo di ricerca delle emozioni [4].

In contrapposizione al filone antropologico, secondo cui le emozioni sono dettate dalla cultura e dalle tradizioni del luogo, Ekman propone la teoria di universalità delle espressioni facciali e delle emozioni stesse, detta anche teoria neuro culturale, sostenuta e scientificamente provata da molteplici esperimenti sul campo, condotti dall'autore e dai suoi collaboratori nel corso degli anni '60-'70.

Il rigoroso approccio scientifico e sperimentale adottato dallo psicologo, riprese gli studi di Darwin sulle espressioni facciali delle emozioni e ne dimostrò la loro universalità, totalmente indipendenti dalla personalità, dall'ambiente e dal contesto di crescita del

soggetto. Egli, dunque, sostiene che esistono espressioni facciali derivanti da emozioni esperite in determinate situazioni, caratterizzate da mimiche universali [5]. Per avvalorare la sua teoria, condusse molteplici esperimenti a soggetti con bagagli culturali differenti e provenienti da diverse parti del mondo.

I risultati ottenuti confermarono quanto sostenuto da Ekman, ovvero l'esistenza di emozioni comuni definite primarie ed elencate in seguito:

- Felicità
- Sorpresa
- Tristezza
- Rabbia
- Paura
- Disgusto

Successivamente, la lista delle emozioni universalmente riconosciute si ampliò ad una sottocategoria composta da ulteriori stati emozionali, che non sono altro che il risultato del mescolarsi di quelle primarie. Tra queste, chiamate secondarie, si annoverano emozioni quali divertimento, eccitazione, imbarazzo, vergogna, soddisfazione e tante altre, date dall'unione di quelle primarie.

In conclusione, si può confermare che gli studi di Ekman sono fonte di conoscenza ed ispirazione, che apportano prove inconfutabili sull'universalità delle emozioni umane.

2.2. PLUTCHIK: LA RUOTA DELLE EMOZIONI

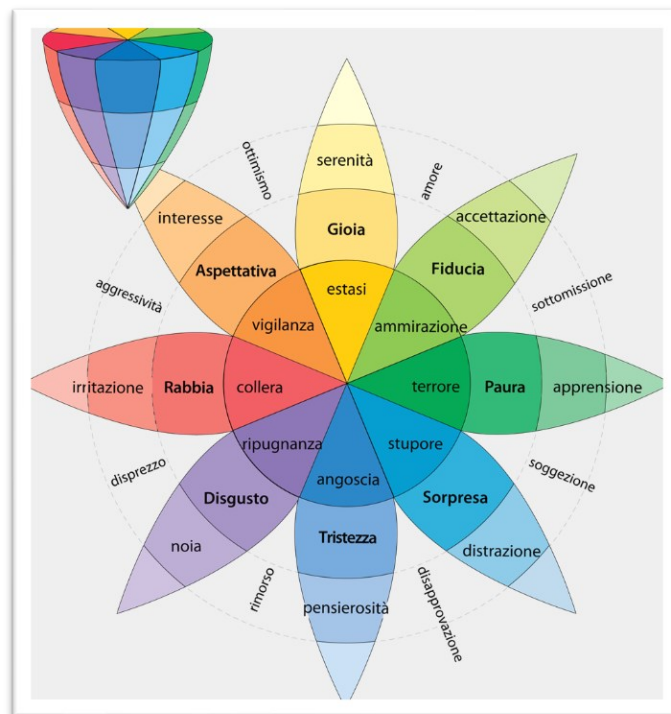
Le emozioni vengono definite come delle reazioni determinate da esperienze, piacevoli e spiacevoli, intense ed immediate, che generano delle percezioni forti in grado di provocare uno stato di alterazione temporaneo dell'equilibrio psicofisico [6].

Nonostante esse siano legate ad alcuni meccanismi molto complessi del sistema neuronale umano, dove la scienza ancora oggi non ha risposte certe, possiamo comunque

contare su innumerevoli studi condotti da psicologi e scienziati, dalla quale sono state prodotte teorie di assoluto valore sia sul loro riconoscimento, che sul loro manifestarsi.

Il noto psicologo statunitense Robert Plutchik rappresenta una delle figure di maggiore rilievo in questo ambito di applicazione, dove nel 1980 sviluppò un modello in grado di fornire una logica alle emozioni e al loro insorgere, chiamato “ruota delle emozioni” e consultabile in figura 1.

Figura 1: La ruota delle emozioni



Fonte: Plutchik, 1980

Nel suo modello, Plutchik (in contrapposizione ad altri autori) ritiene che le emozioni primarie siano otto, proprie di ogni individuo e facilmente identificabili secondo le diverse colorazioni in figura 1. Inoltre, combinate assieme ne generano di altre e più complesse, chiamate emozioni secondarie.

L'idea alla base di tale teoria è che esiste una differenza sostanziale tra primarie e secondarie: le prime sono innate nell'individuo a livello biologico, mentre le seconde sono da ritenersi come un'evoluzione di quelle primarie [7].

Secondo tale autore, le emozioni primarie sono:

- Gioia
- Tristezza
- Rabbia
- Paura
- Fiducia
- Disgusto
- Sorpresa
- Anticipazione

Quest'ultime, mescolate tra loro generano le emozioni secondarie quali amore, ottimismo, aggressività, disprezzo e tante altre.

La ruota delle emozioni, inoltre, fornisce maggiori informazioni a riguardo:

- Le emozioni primarie sono disposte in settori e ogni coppia è disposta in modo tale che a ogni emozione corrisponda il suo opposto;
- Ogni settore ha una colorazione sfumata proprio per evidenziare l'intensità dell'emozione stessa;
- Le zone di interesse tra uno settore e l'altro (zone bianche in figura) rappresentano le emozioni secondarie;

In sintesi, il modello proposto da Plutchik offre un approccio categorico all'identificazione delle emozioni, dettagliato inoltre dall'intensità del colore che ne differenzia i vari strati emotivi e le relazioni tra essi.

La ruota delle emozioni è considerata quindi uno strumento esplicativo in grado di interpretare correttamente il contesto emotivo oggetto di studio. Essere consapevoli dello stato emotivo in cui ci si trova, permette ai soggetti di prestare più attenzione ai processi decisionali, considerando quindi un maggior numero di fattori rilevanti e le conseguenze che ne derivano da essi.

3. OBIETTIVI DI ANALISI

Le piattaforme digitali sono note per offrire la possibilità agli utenti di creare e condividere contenuti multimediali di loro interesse con altri soggetti appartenenti al network. Tali contenuti possono essere di varie tipologie, anche a seconda del tipo di social network utilizzato, ma la forma di comunicazione più diffusa e onnipresente in qualsiasi piattaforma è il formato testuale. Sebbene esistano social che prediligono mezzi di comunicazione alternativi, come immagini o video, il testo risulta ancora oggi un elemento fondamentale per tutti coloro che vogliono interagire attivamente all'interno della community ed esprimere opinioni, idee, sentimenti, emozioni e quant'altro.

In quest'ottica, Twitter rappresenta il social network testuale per antonomasia, dove chiunque può pubblicare contenuti testuali e condividere quelli di altri. Esso è un servizio di microblogging ampiamente diffuso negli Stati Uniti e non solo, utilizzato soprattutto per diffondere notizie real-time, soprannominato per l'appunto l'SMS di internet [9]. Permette non solo di comunicare, ma soprattutto di esprimere sentimenti ed emozioni tramite breve testo scritto, chiamato tweet.

L'utilizzo che se ne fa di Twitter nell'accezione più comune, ha dato origine a ricerche e analisi di contenuti molto interessanti, dove secondo studi recenti gli utenti fanno ampio uso della piattaforma per esprimere un proprio consenso o dissenso nei confronti di un determinato argomento, mostrando apertamente il proprio stato emotivo. Proprio quest'ultimo è stato oggetto di studi scientifici in quanto la sua esatta determinazione permette di estrapolare informazioni preziose in molti campi di applicazione.

Sebbene esistano servizi tecnologici avanzati di determinazione del sentiment in grado di comprendere se l'autore del contenuto (testuale o altro) esprima un senso di approvazione o disapprovazione, più rare sono invece le applicazioni in grado di riconoscere l'emozione predominante che governa il contenuto pubblicato dall'utente.

Per tale ragione, l'attuale lavoro focalizza l'attenzione sul riconoscimento delle emozioni espresse dagli utenti mediante contenuti testuali, attività che nel gergo tecnico viene definita di emotion recognition.

L'emotion recognition è definito come un processo di identificazione delle emozioni umane che avviene mediante tecnologie digitali, basato su costrutti in grado di integrare e combinare studi psicologici con calcolo computazionale [10].

Al giorno d'oggi, sono stati costruiti diversi classificatori di emozioni e set di dati per i testi in lingua inglese. Tuttavia, nonostante un gran numero di applicazioni di riconoscimento delle emozioni nei testi, solo poche risorse lessicali e modelli di classificazione sono disponibili per altre lingue [11].

Il progetto, quindi, ha come obiettivo principale lo sviluppo di modelli per le attività di emotion recognition, utilizzando contenuti testuali provenienti dai social e scritti in lingua italiana.

I modelli sviluppati e confrontati tra loro, sapranno predire l'emozione che predomina lo stato emotivo dell'utente nel momento in cui scrive un tweet, un post, un commento o usufruisca di un qualsiasi altro input testuale.

Per poter realizzare con successo l'obiettivo preposto, si fa ampio utilizzo durante il lavoro dei principali algoritmi di deep learning, ovvero reti neurali idealmente progettate per assolvere problemi specifici, che rientrano in diversi campi di applicazione, tra cui l'elaborazione del linguaggio naturale.

4. TECNOLOGIE UTILIZZATE

Il seguente capitolo è pensato per descrivere le principali architetture e tecnologie utilizzate durante le fasi di test del progetto. Seguirà quindi una panoramica dei contributi offerti dai sistemi sempre più avanzati di intelligenza artificiale, come machine learning, deep learning e i loro utilizzi nello specifico ambito di elaborazione del linguaggio naturale.

Particolare attenzione poi, sarà dedicata al funzionamento delle singole tecnologie dal punto di vista tecnico, nonché al motivo per le quali sono state implementate e le soluzioni da loro proposte per il raggiungimento di un determinato obiettivo.

4.1. ARTIFICIAL INTELLIGENCE

L'Artificial Intelligence (AI), è il ramo della computer science che studia lo sviluppo di sistemi Hardware e Software dotati di specifiche capacità tipiche dell'essere umano, capaci di perseguire autonomamente una finalità definita, prendendo decisioni che fino a quel momento erano solitamente affidate alle persone [12].

Dalla definizione si può comprendere come l'obiettivo dell'intelligenza artificiale sia quello di replicare le capacità cognitive tipiche dell'essere umano, sfruttando potenza computazionale e unità di calcolo messe a disposizione dai sistemi moderni.

Le applicazioni di intelligenza artificiale fanno largamente uso di algoritmi e tecniche computazionali in grado di replicare il comportamento degli esseri umani e le proprietà cognitive di cui sono dotati, agendo così da sistemi razionali in grado di percepire l'ambiente circostante e mettere in atto azioni su di esso.

Sebbene il concetto di intelligenza artificiale fu coniato 70 anni fa, nel corso della storia il suo sviluppo ed il suo utilizzo subì notevoli rallentamenti dovuti principalmente da limitazioni di tipo hardware. Se i modelli matematici diventarono sempre più sofisticati ed in grado di imitare alcune delle funzionalità tipiche dell'essere umano, la limitata potenza computazionale, le architetture hardware e l'effettiva mancanza di dati disponibili e facilmente accessibili, non permisero loro di crescere a dovere.

La maturità tecnologica raggiunta negli anni '90, grazie a soluzioni hardware molto potenti e di ridotte dimensioni, permisero di propagare le ricerche in questo campo di applicazione e, grazie anche all'arrivo dei Big Data, esplose il fermento attorno a questa disciplina [13].

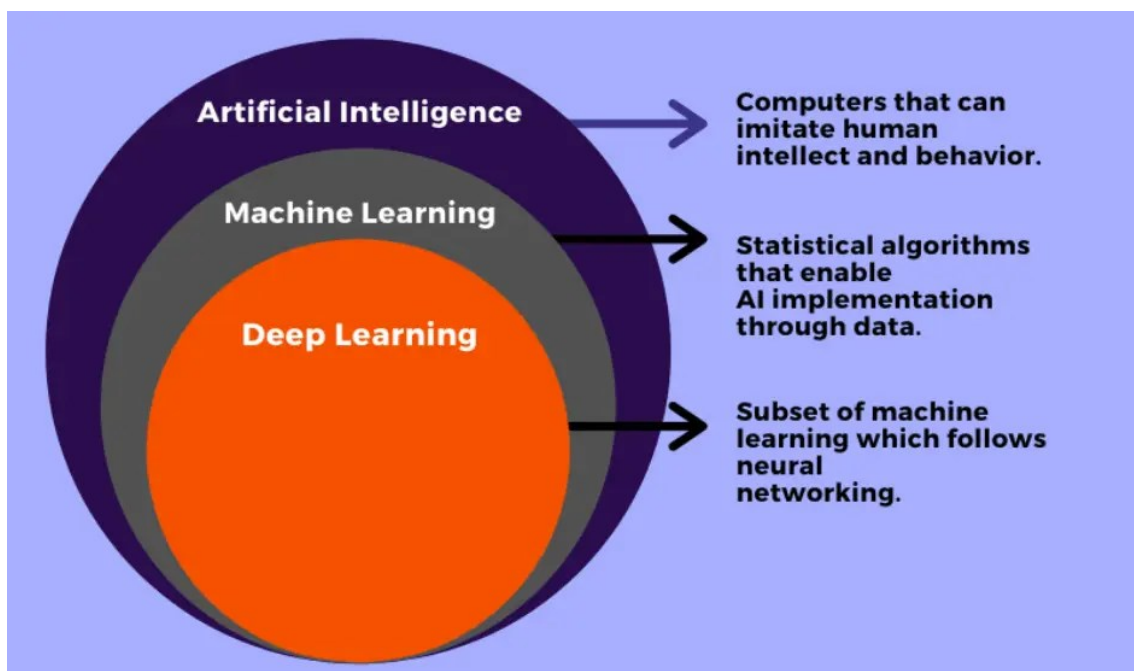
È proprio grazie alla grande mole di dati reperibili in rete che iniziarono a diffondersi sistemi basati su intelligenza artificiale in grado di elaborare i dati in ingresso, apprendere automaticamente pattern presenti in essi e infine, restituire un risultato di output.

In quest'ottica, è doveroso introdurre i concetti di machine learning e deep learning che verranno discussi nei prossimi paragrafi, in modo tale da trattare in maniera esaustiva tutte le parti componenti di un'intelligenza artificiale moderna.

4.1.1. Machine learning

Machine learning, o apprendimento automatico, è un termine coniato nel 1959 da Arthur Samuel, con la quale indicò una raccolta di metodi che consentono ai computer di apprendere senza essere esplicitamente programmati dall'uomo [14].

Figura 2: Artificial Intelligence, Machine Learning and Deep Learning



Fonte: AI4Business, 2020

Come si nota in figura 2, sia machine learning che deep learning sono due sottoinsiemi dell'intelligenza artificiale.

Il machine learning in particolare consiste in specifici algoritmi matematici-statistici in grado di apprendere informazioni direttamente dai dati e migliorare le loro prestazioni in maniera adattiva, mediante un apposito modello di apprendimento. Proprio quest'ultimo permette all'AI di imparare dall'esperienza, allenandola in modo tale da svolgere autonomamente specifiche attività e compiti ben precisi. In questi casi, si dice che il

sistema apprende dall'esperienza quando le sue prestazioni migliorano dopo lo svolgimento di un compito o il completamento di una specifica azione [15].

È nota nel mondo accademico una definizione moderna di machine learning offerta dall'ex presidente del dipartimento di machine learning presso la Carnegie Mellon University, Tom Mitchell:

si dice che un programma per computer apprenda dall'esperienza E rispetto ad una classe di compiti T , le cui prestazioni sono misurate da P , se le sue prestazioni nei compiti T , misurate da P , migliorano con l'esperienza E [16].

in sintesi, tali algoritmi modificano le performance di un sistema a seconda della quantità e qualità dei dati che gli vengono forniti in input per allenarsi. Quindi nell'attuale era dei Big Data essi hanno assunto un ruolo centrale in molti ambiti di applicazione per via dei numerosi vantaggi che apportano a coloro che gli implementano.

4.1.2. Deep learning

Con il termine deep learning, ci si riferisce ad un'ampia sottocategoria del machine learning, composta da un insieme di tecniche basate su reti neurali artificiali organizzate in diversi strati ed ispirate al funzionamento biologico dei neuroni [17].

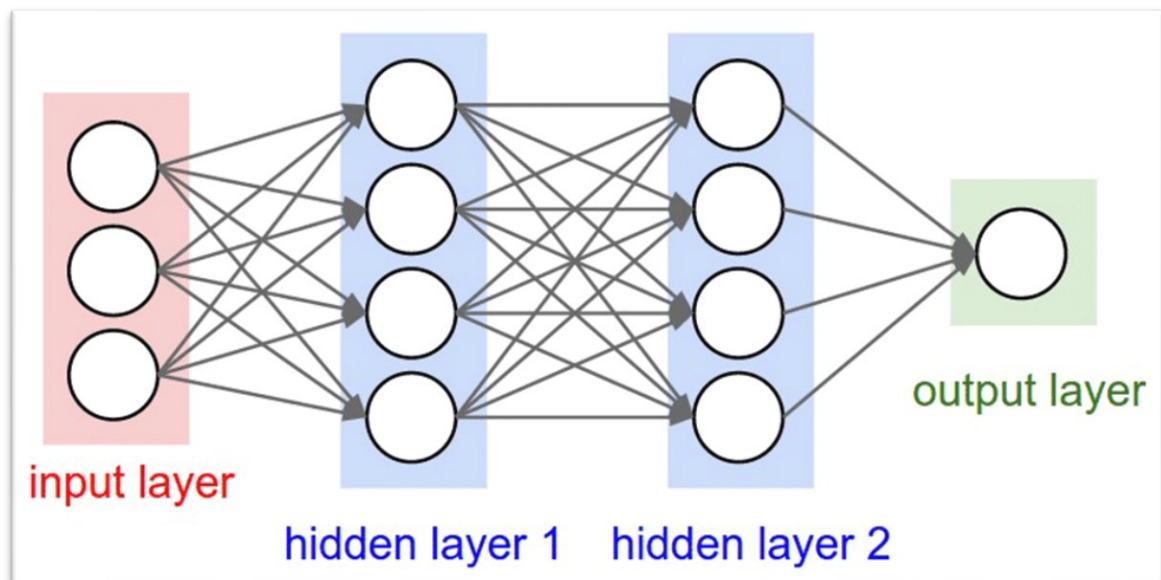
Il deep learning è quindi una branca dell'intelligenza artificiale che sfrutta le ottime prestazioni delle reti neurali in un'ampia gamma di applicazioni. Tra queste, gli ambiti applicativi più comuni sono elencati in seguito:

- Classificazione delle immagini
- Riconoscimento ed elaborazione del linguaggio naturale
- Sicurezza (es. videosorveglianza e riconoscimento facciale)
- Diagnosi mediche
- Sistemi di raccomandazione
- Guida autonoma (es. rilevamento pedoni, segnali stradali)

- Identificazione di frodi

Come facilmente intuibile, i capisaldi di un sistema di deep learning sono le reti neurali artificiali, ovvero dei complessi modelli computazionali composti da diversi strati di neuroni, interconnessi tra loro per lo scambio e l'elaborazione di informazioni. Nella maggior parte dei casi una rete neurale artificiale è un sistema adattivo che cambia la propria struttura in base a informazioni esterne o interne che scorrono attraverso la rete stessa durante la fase di apprendimento [18].

Figura 3: Structure of Artificial Neural Network



Fonte: NetAi.it, 2022

Sebbene esistano numerose tipologie di reti neurali, dal più antico Perceptrons fino alle più moderne Attention Network, è possibile analizzare brevemente i tre strati comuni presenti nella maggior parte delle reti deep, visibili in figura 3:

- Input layer: strato in cui sono presenti i neuroni di input, caratterizzati dai dati in ingresso
- Hidden layer: uno o più livelli nascosti situati tra lo strato di input e lo strato di output, in cui sono presenti i neuroni incaricati di eseguire i calcoli computazionali

- Output layer: è il livello di output che produce i risultati finali

Sempre dall'immagine presente in figura 3, si può osservare come ogni neurone appartenente ad uno strato è connesso ad ogni neurone presente nello strato successivo e ogni connessione ha un peso specifico particolare, dato dall'importanza che l'informazione trasmessa dal neurone ha nel processo di apprendimento [19].

In seguito, verrà illustrato il funzionamento di una particolare rete neurale oggetto di studio del progetto, considerata tra le più illuminanti ed innovative per applicazioni legate all'elaborazione del linguaggio naturale.

4.1.3. Natural Language Processing

Il Natural Language Processing (NLP), o elaborazione del linguaggio naturale, è definito come un ramo dell'intelligenza artificiale che si occupa dell'interazione tra Data Science e linguaggio umano, in grado di fornire ai computer la capacità di comprendere il testo e le parole pronunciate in linguaggio naturale. In tal senso lo scopo è quello di rendere le macchine capaci di leggere, comprendere e ricavare significato dai linguaggi tipici dell'uomo [20].

Tra le problematiche più grandi riscontrate in questa disciplina, vi è senz'altro la difficoltà da parte di un algoritmo di interpretare correttamente le numerose ambiguità presenti in un discorso tra esseri umani, il che complica notevolmente la comprensione del messaggio. Tutto ciò, può condurre a comunicazioni contraddittorie e prive di significato [21].

In passato le problematiche relative alle ambiguità del linguaggio si tentavano di superare mediante un approccio sequenziale composto da fasi che si succedevano l'una con l'altra.

Negli ultimi anni si è assistito alla nascita di nuovi approcci, che integrano l'elaborazione del linguaggio naturale con gli algoritmi di deep learning, ottenendo risultati nettamente superiori rispetto al passato. Si è dimostrato come le deep neural network siano in grado

di produrre risultati considerevoli se applicate a specifici task di Natural Language Processing.

Tra le tecniche più popolari in ambito deep learning, i word embeddings rappresentano sicuramente una delle tecnologie più avanzate per l'analisi del linguaggio naturale.

Essi sono definiti come una classe di tecniche in cui le singole parole sono mappate a vettori numerici in un determinato spazio vettoriale. Il vantaggio principale di queste tecnologie riguarda il posizionamento del vettore nello spazio: la sua distribuzione viene appresa in base all'uso delle parole, consentendo a quest'ultime di avere rappresentazioni simili se utilizzate in contesti simili, catturando così il loro significato semantico [22].

Sebbene le tecniche appena descritte siano ottime, hanno il limite di non tenere in considerazione dell'ordine delle parole nel testo. Per tale motivo sono nati di recente ulteriori modelli in grado di superare tale problema, che verranno descritti nel paragrafo seguente.

4.1.4. Transformers

Un Transformers è definito come un modello di deep learning che fa largo uso del meccanismo dell'attenzione, ponderando in modo differenziale il significato di ciascuna parte dei dati di input [23]. Sono considerati i modelli più all'avanguardia per applicazioni in ambito Natural Language Processing, sviluppati principalmente per risolvere il problema di traduzione automatica neurale, ovvero tutte quelle attività dove è necessario trasformare sequenze di input in sequenze di output.

Nascono per risolvere alcuni problemi che altre architetture neurali non riuscivano a risolvere, come le Recurrent Neural Network (RNN) e le Convolutional Neural Network (CNN).

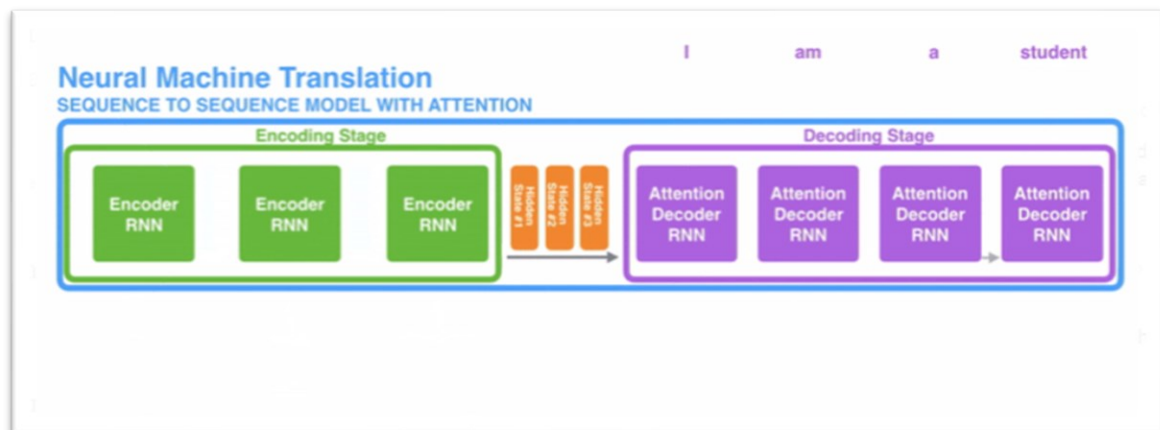
Sebbene le RNN siano studiate per consentire alle informazioni passate di persistere mediante loop interni, diventano inefficaci quando il divario tra le informazioni rilevanti e il punto in cui sono necessarie è ampio.

Le Long-Short Term Memory (LSTM), una particolare architettura di rete neurale ricorrente, ha provato a risolvere tale problema tramite l'utilizzo di celle di memoria, in grado di catturare lo stato di memoria della cella precedente e il suo output, così da passare il tutto alla cella successiva. In questo modo, le LSTM possono ricordare o dimenticare in modo selettivo cose importanti e cose meno importanti [24].

Anche se apparentemente sembrano funzionare, anch'esse perdono di efficacia quando le frasi sono troppo lunghe e il divario tra una sequenza significativa e la sequenza corrente è molto ampio.

I ricercatori svilupparono così una tecnica chiamata "di attenzione", in grado di dotare le reti neurali di capacità di concentrazione, individuando così solo parti importanti di un insieme di parole. Hanno lo scopo di replicare l'attenzione cognitiva tipica dell'essere umano, concentrandosi sull'alta risoluzione in alcune parti degli input, mentre il resto degli input è in bassa risoluzione.

Figura 4: sequence to sequence model with attention



Fonte: Towardsdatascience.com, 2022

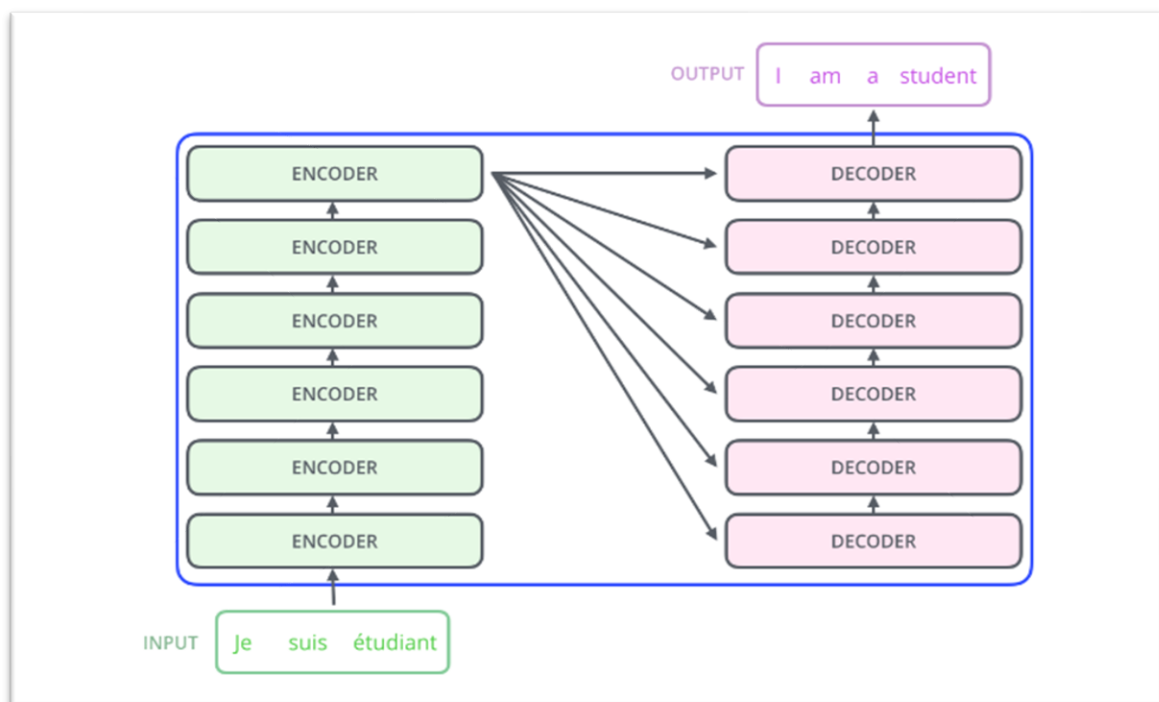
Come si evince in figura 4, la tecnica di attenzione è composta da due fasi, una di codifica (parte verde in figura) e una di decodifica (parte viola in figura). Per le RNN, invece di codificare l'intera frase in un solo strato nascosto, ogni singola parola della frase ha un proprio strato nascosto che viene passato fino alla fase di decodifica, dove le parti

arancioni in figura indicano lo strato nascosto dei vettori di input (tre parole in questo caso). Quindi, tutti gli hidden layers vengono utilizzati in fase di decodifica, per capire dove la rete dovrebbe dedicare maggiore attenzione [25].

Si riscontrano tuttavia alcuni problemi relativi all'implementazione di tecniche di attenzione su RNN, uno tra tutti l'impossibilità da parte della rete di elaborare le parole in modo parallelo e ciò, causa un problema quando il corpus di testo è relativamente grande.

Si introducono a questo punto i famigerati Transformers, un'architettura particolare in grado di risolvere i problemi precedentemente enumerati, mediante l'utilizzo di codificatori e decodificatori. Al suo interno, un Transformers ha un'architettura simile a quella dei modelli precedenti, ma con più unità di codificatori e decodificatori, come visibile in figura 5.

Figura 5: Architecture of Transformers

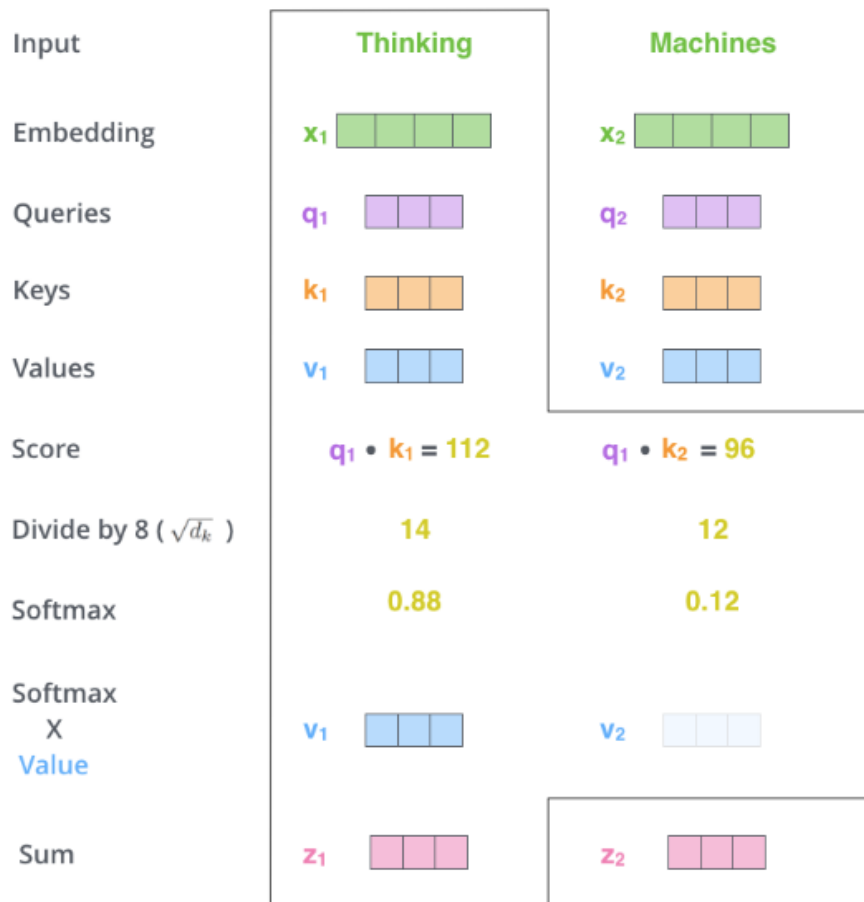


Fonte: Towarddatascience.com, 2022

Ogni encoder ha al suo interno un livello di auto attenzione e un livello composto da una rete neurale feed forward. Un decoder ha invece un livello in più di attenzione rispetto agli encoder, che aiuta il decodificatore a concentrarsi su parti rilevanti della frase di input.

Nonostante i transformers siano architetture complesse composte da parti componenti collegate ed interconnesse tra loro, il fulcro centrale che garantisce loro di funzionare a dovere è caratterizzato dal meccanismo di auto attenzione, il cui processo è visibile in figura 6.

Figura 6: Self-attention mechanism



Fonte: Towardsdatascience.com, 2022

Il primo passaggio consiste nel creare gli incorporamenti di parole di input del codificatore, in vettori di dimensione pari a 512. Successivamente verranno creati tre vettori: vettore Query, vettore Keys e vettore Values, tutti di dimensione pari a 64. Questi rappresentano un'astrazione di alto livello utile per calcolare l'attenzione, ottenuti moltiplicando gli embeddings per tre matrici addestrate durante il processo di training.

Successivamente verrà calcolato un punteggio in grado di determinare quanta attenzione dedicare alle altre parole di input quando codifichiamo la parola di interesse in una determinata posizione. Il punteggio viene ricavato facendo il prodotto scalare del vettore Query per il vettore Keys.

Il terzo e il quarto passaggio consistono nel dividere il punteggio per 8 (la radice quadrata della dimensione dei vettori Query, Keys e Value) e passare il risultato ad una funzione Softmax in grado di normalizzare il punteggio in un range da 0 a 1.

Il quinto passaggio consiste nel moltiplicare il punteggio Softmax per il vettore Values, così da mantenere intatti i valori delle parole su cui si vuole dedicare attenzione e, al contempo, di nascondere quelle irrilevanti.

Il sesto ed ultimo passaggio, consiste nel sommare i vettori dei valori ponderati, così da produrre l'output dello strato di auto attenzione in questa posizione.

Tale risultato è l'output che verrà inviato alla fase successiva, cioè alla rete feed forward [24].

La descrizione tecnica riportata circa i Transformers e i loro articolati meccanismi di funzionamento, sono utili per comprendere le tecnologie quali GPT-3 e BERT testate nelle fasi sperimentali dell'attuale progetto, in quanto quest'ultime non sono altro che trasformatori che devono le loro alte performance ai meccanismi appena descritti.

4.2. UNSUPERVISED LEARNING

Come noto in letteratura, l'unsupervised learning, o apprendimento non supervisionato, identifica una particolare classe di algoritmi in grado di riconoscere pattern e relazioni esistenti nei dati, senza che quest'ultimi abbiano un corrispondente valore di output.

Il suo utilizzo è adatto a cercare modelli nascosti nei dati che sfuggono all'osservazione, sia perché oscurati da altre informazioni sia perché la quantità di dati è talmente grande da non poter essere osservata facilmente senza un ausilio computazionale [26].

Nel seguente progetto, sono stati implementati algoritmi di tipo unsupervised per diverse attività esplorative e di scoperta di relazioni nei dati. Una volta creati gli incorporamenti di parole (word embeddings), le tecniche non supervisionate hanno permesso di fare analisi cluster per scovare pattern interni e nascosti.

Nei prossimi paragrafi verranno descritti due algoritmi utilizzati in questa fase, ovvero Word2Vec per incorporamenti di parole e K-means per attività di clustering.

4.2.1. Word2Vec

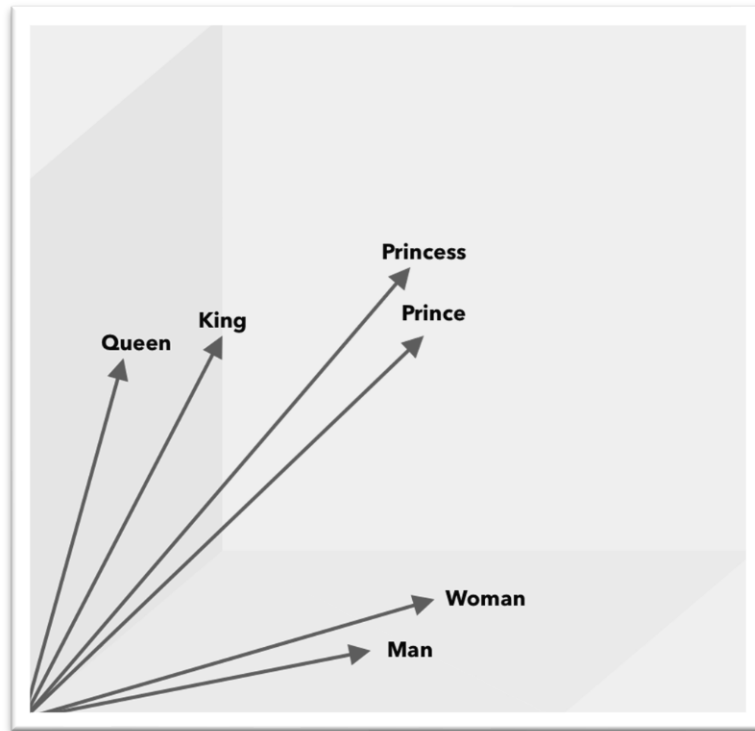
Word2Vec è definito come un insieme di modelli in grado di produrre incorporamenti efficienti di parole, utilizzando una rete neurale semplice a due strati che prende in input un corpus di testo e produce, in output, una rappresentazione vettoriale in grado di catturare la distribuzione semantica delle parole presenti nel testo.

Tale tecnica è stata ampiamente utilizzata in ambito Natural Language Processing perché apprende associazioni tra parole utilizzando una funzione matematica in grado di identificare il livello di somiglianza semantica.

L'obiettivo dell'algoritmo è quello di creare word embeddings, ovvero riprodurre le parole in uno spazio vettoriale di diverse centinaia di dimensioni, con ogni parola univoca

nel testo a cui viene assegnato un vettore corrispondente nello spazio. I vettori di parole sono posizionati nello spazio vettoriale in modo tale che le parole che condividono contesti comuni nel corpus si trovano vicine l'una all'altra nello spazio [27].

Figura 7: Word embeddings with Word2Vec algorithm

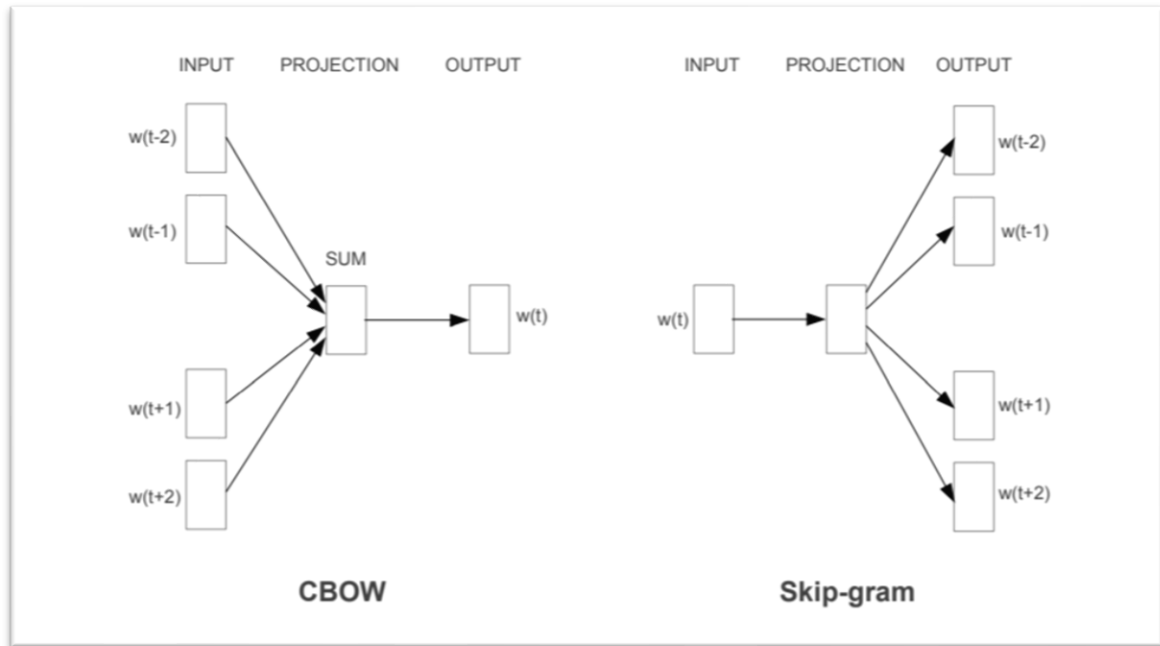


Fonte: Towardsdatascience, 2020

Osservando la figura 7, è intuibile come parole con significato simile si verificano spazialmente nelle immediate vicinanze, matematicamente misurata dalla somiglianza del coseno, dove la somiglianza totale è espressa come un angolo di zero gradi, mentre nessuna somiglianza è espressa come un angolo di novanta gradi [28].

Sebbene le parole possano essere codificate utilizzando diversi tipi di modelli quali reti neurali feed forward e reti neurali ricorrenti, i risultati migliori sono stati ottenuti implementando architetture log-lineari, quali Continuous Bag Of Word (CBOW) e Skip Gram, visibili in figura 8.

Figura 8: CBOW vs Skip-Gram



Fonte: *Efficient Estimation of Word Representation in Vector Space*, 2013

Nell'architettura CBOW il modello prevede la parola corrente da una finestra di parole di contesto circostante, ed è per un certo senso simile al funzionamento di una rete neurale feed forward.

Mikolov, autore e creatore dell'algoritmo Word2Vec, afferma che la migliore prestazione mediante l'utilizzo dell'architettura CBOW si è verificata utilizzando una finestra di contesto che prende le quattro parole future e le quattro parole precedenti rispetto alla parola target. Inoltre, è assai importante considerare il fatto che l'ordine delle parole di contesto non influenzano i risultati dell'output.

Al contrario, Skip Gram funziona esattamente l'opposto di CBOW, ovvero il modello utilizza la parola corrente per prevedere la finestra circostante delle parole di contesto. In questa architettura, la parola corrente è l'input che viene passato al modello e l'obiettivo è quello di prevedere con precisione le parole prima e dopo la parola di input. A differenza di CBOW, questo approccio considera le parole di contesto vicine più pesantemente delle parole di contesto lontane.

La qualità dei word embeddings ottenuti mediante l'algoritmo Word2Vec è sensibile a parametrizzazione. La dimensionalità è un parametro assai importante, maggiore è la dimensione dello spazio vettoriale, maggiore è la qualità degli incorporamenti di parole. Dopo però aver raggiunto una certa soglia, il guadagno marginale diminuisce, così gli autori consigliano di impostare tale parametro in un range da 100 a 1000 [27].

Anche il parametro window, ovvero la finestra di contesto che determina quante parole prima e dopo la parola corrente devono essere considerate nel modello, è in grado di influenzare sensibilmente i risultati finali. Gli autori consigliano un valore pari a 10 se si utilizza un'architettura Skip Gram, mentre un valore pari a 5 per CBOW.

Infine, l'algoritmo di addestramento influisce sui risultati finali. Un modello Word2Vec può essere addestrato mediante due categorie di algoritmi, softmax gerarchico e campionamento negativo.

Ciò che occorre sapere per l'impostazione di tale parametro è che, sempre secondo Mikolov, softmax gerarchico funziona meglio per parole poco frequenti mentre il campionamento negativo garantisce maggior qualità per parole frequenti e con uno spazio vettoriale modesto.

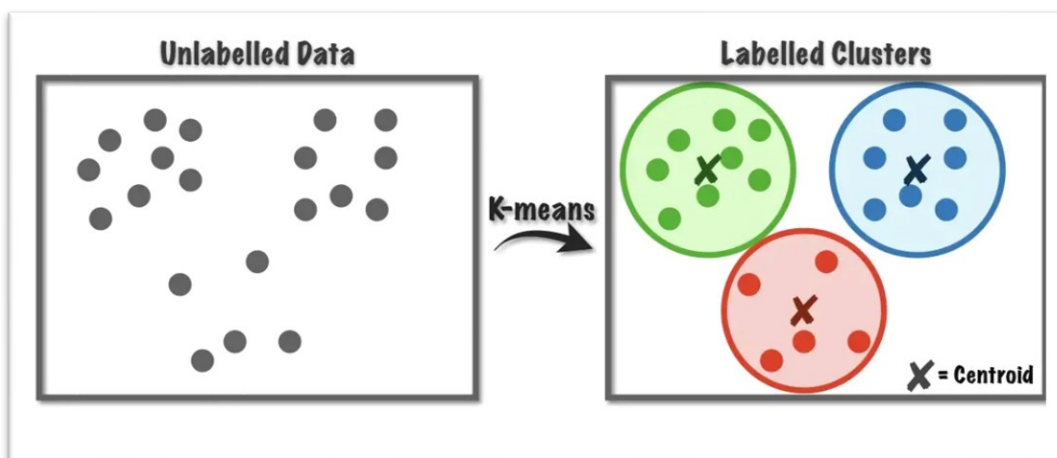
Nell'attuale lavoro, Word2Vec è stato utilizzato per creare word embeddings in grado di rappresentare tutte le parole del corpus di testo sotto forma di vettori. Sono state testate diverse configurazioni di Word2Vec, lasciando la descrizione dettagliata del parameter tuning al capitolo 4.

Una volta ottenuti gli incorporamenti, sono stati passati come input ad algoritmi di clustering al fine di determinare quali pattern potevano esserci tra le parole utilizzate in contesti differenti.

4.2.2. K-means

K-means è un algoritmo di apprendimento automatico unsupervised progettato per fare analisi cluster, in grado di suddividere i dati non etichettati in gruppi omogenei, la cui omogeneità è misurata in base agli attributi dei dati stessi, ovvero alle loro caratteristiche.

Figura 9: K-means algorithm



Fonte: Towardsdatascience.com, 2019

Il suo funzionamento è piuttosto semplice, permette di raggruppare i dati simili in modo tale da rendere visibile eventuali pattern o relazioni tra gli stessi che in assenza di raggruppamento non sarebbe stato facile coglierli.

L'algoritmo segue una procedura iterativa, il cui obiettivo è quello di minimizzare la varianza totale intra-gruppo e massimizzare quella inter-gruppo, in modo tale da avere dati simili nello stesso cluster e, al contempo, dati dissimili in cluster diversi.

Viene scelto in primo luogo il valore di K che rappresenta il numero di cluster finali che l'algoritmo restituisce, poi in modo casuale un centroide iniziale viene definito in un punto qualsiasi del piano, e rappresenta le coordinate centrali di un cluster. Si ottengono così K centroidi, ognuno rappresentanti il centro di un cluster.

In maniera iterativa si succederanno due fasi [29]:

- Fase di assegnazione, dove ogni osservazione viene assegnata al centroide più vicino;
- Fase di aggiornamento, dove i centroidi vengono aggiornati a seconda delle osservazioni assegnate ai vari cluster;

questi passaggi vengono ripetuti più e più volte, finché nessuna osservazione cambierà di cluster e quindi le coordinate dei centroidi non verranno aggiornate ma si fermeranno nella posizione corrente.

Il seguente lavoro prevede l'implementazione dell'algoritmo K-means con il fine di raggruppare gli incorporamenti di parole ottenuti mediante Word2Vec. In tal modo, si cerca di mostrare quali legami esistono tra le parole usate in frasi diverse e, facendo opportunatamente uso dell'etichetta assegnata alla frase, si valuta se c'è corrispondenza tra vicinanza semantica e vicinanza emotiva della parola.

4.3. SUPERVISED LEARNING

In contrapposizione all'apprendimento non supervisionato descritto in precedenza, il supervised learning si riferisce a modalità di apprendimento che utilizzano dati di addestramento già etichettati in classi predefinite, garantendo all'algoritmo di imparare a prevedere correttamente le classi di nuovi oggetti di dati non etichettati.

In un contesto di tipo supervisionato, l'algoritmo permette di apprendere le correlazioni comuni tra i dati appartenenti alla stessa classe e l'obiettivo consiste nell'effettuare previsioni corrette di dati nuovi non etichettati. Ad ogni dato di input si conosce a priori il rispettivo output che viene utilizzato per addestrare il modello in diverse iterazioni.

Tale modalità di apprendimento è più laboriosa e complessa rispetto l'unsupervised learning, perché prevede la preparazione sia di un set di dati di training che garantisce la presenza dell'etichetta nei dati, sia di un set di dati di test su cui poi si andrà a testare le

prestazioni del modello. L'algoritmo viene quindi allenato sul dataset di addestramento producendo un modello che verrà provato sul test set, composto da dati non etichettati.

Nel seguente progetto, grande attenzione è stata dedicata alla parte supervisionata con il fine ultimo di assolvere all'obiettivo principale del lavoro, ovvero la creazione di un modello di classificazione in grado di riconoscere le emozioni umane nel testo.

Nei prossimi paragrafi verranno descritti due algoritmi utilizzati in questa fase sperimentale, che hanno permesso lo sviluppo di due differenti modelli di classificazione.

4.3.1. GPT-3

Generative Pre-trained Transformer 3, meglio noto come GPT-3, è un modello linguistico creato da OpenAI (laboratorio scientifico di intelligenza artificiale con sede a San Francisco) e ora di proprietà della Microsoft corporation, in grado di produrre testo simile al linguaggio umano e considerato il più grande sistema linguistico mai creato in precedenza [30].

L'architettura di GPT-3 ha 175 miliardi di parametri dislocati in 96 livelli, addestrati su 500 miliardi di dati di contenuti web. Per addestrare il modello, la maggior parte dei dati sono stati acquisiti da una versione filtrata del common crawl, un archivio internet che contiene circa 25 petabyte di dati, e la restante parte da libri, articoli e post online, oltre all'intero corpus di informazioni contenute in Wikipedia [31].

GPT-3 è in grado di assolvere diverse funzioni linguistiche in quanto utilizza algoritmi di deep learning pre-addestrati, cioè che hanno già immagazzinato tutti i dati necessari per poter svolgere il compito richiesto dall'utente.

La principale funzione di GPT-3 è la comprensione e l'elaborazione del linguaggio naturale, dopodiché svolge attività di completamento del testo, scrittura di testi, articoli e poesie, scrittura di codice, risposta a domande specifiche, classificazione di documenti e tanto altro.

GPT-3 apprende in modo non supervisionato ottenendo le informazioni necessarie al completamento di un task dall'analisi dei dati contenuti nel suo enorme database.

Inizialmente commette errori, ma man mano che il processo di ricerca continua troverà la risposta corretta verificando i dati di input originariamente passati al modello per l'addestramento e assegnerà un peso all'intero processo che ha portato a quella soluzione [32].

Il codice sorgente di GPT-3 non è reso disponibile dagli sviluppatori, ma permettono comunque di utilizzare Application Programming Interface (API) per interfacciarsi con il modello e sfruttarlo per qualsiasi tipo di richiesta, purché abbia a che fare con testo. È quindi un servizio privato non open source a pagamento il cui costo varia in base all'attività da svolgere e al modello di fine tuning scelto dall'utente.

Il seguente lavoro ha tratto profitto dalle enormi capacità di GPT-3 per mettere a punto un modello di classificazione per riconoscere le emozioni nel testo, testando sia i motori standard che il motore più performante in grado di offrire la soluzione migliore al problema. Infine, sono stati comparati i risultati ottenuti da GPT-3 con quelli ottenuti da un altro modello linguistico altamente performante, descritto nel paragrafo successivo.

4.3.2. Google BERT

BERT è l'acronimo di Bidirectional Encoder Representation from Transformers e consiste in un modello di deep learning progettato da Google per elaborare il linguaggio naturale.

In particolare, BERT utilizza la tecnologia Transformers già descritta in precedenza, in modo da cogliere le relazioni contestuali tra le parole in un testo e fornire un risultato previsionale più accurato.

A differenza dei modelli direzionali che leggono l'input in sequenza da sinistra a destra (o viceversa), il codificatore Transformers legge l'intera sequenza di parole

contemporaneamente consentendo così al modello di apprendere il contesto di una parola da ciò che lo circonda [33]. In questo modo BERT acquisisce un senso più profondo del flusso linguistico rispetto ai tradizionali modelli unidirezionali.

Così come in GPT-3, anche in BERT il modello viene prima pre-addestrato mediante una rete neurale molto grande, il cui costo computazionale è assai oneroso, per produrre incorporamenti di parole da utilizzare come funzionalità nei modelli NLP.

L'architettura di BERT prevede l'utilizzo di una rete neurale a più livelli di encoder che ha lo scopo di trasformare il testo di input in vettori numerici, embedding appunto. Viene poi sfruttato il meccanismo dell'auto attenzione (già ampiamente discusso nei paragrafi precedenti) per calcolare il livello di quest'ultimo sull'embedding delle parole al livello precedente, con lo scopo di creare nuovi incorporamenti intermedi.

Il modello base di BERT ha 12 strati intermedi, produce quindi 12 rappresentazioni intermedie per lo stesso token mentre il modello large ha 24 livelli intermedi e produce altrettante 24 diverse rappresentazioni intermedie per lo stesso token.

Le ottime prestazioni offerte dal seguente modello sono rese possibili grazie ad una attenta fase di pre-training in cui, oltre a utilizzare enormi quantità di dati e tanta potenza computazionale, si mettono in atto due strategie di formazione:

- Masked Language Model, dove il 15% delle parole di una sequenza vengono sostituite con un token MASK e BERT deve effettuare previsioni sul valore originale del token in base al contesto fornito da parole non mascherate;
- Next Sentence Prediction, in cui il modello riceve coppie di frasi come input e deve cercare di prevedere se la seconda frase è o meno la frase successiva alla prima [34];

Una volta finita la fase di addestramento, il modello può essere richiamato mediante apposite librerie e messo a punto anche con meno risorse computazionali su set di dati più piccoli e ottimizzando le prestazioni per task specifici.

Nel dettaglio, il progetto ha messo a punto un modello BERT a 24 livelli per creare un classificatore ottimizzato a riconoscere le emozioni nel testo, così come per GPT-3, e infine comparato con quest'ultimo mediante l'utilizzo di alcune metriche di valutazione per vedere quale delle due tecnologie rispondesse meglio alla specifica attività di emotion recognition.

Nel capitolo successivo verrà illustrato dettagliatamente il set-up sperimentale messo a punto per implementare tutte le tecnologie ampiamente illustrate in questa fase.

5. SET-UP SPERIMENTALE

Il presente capitolo ha lo scopo di mostrare le implementazioni di tutte le relative tecnologie discusse in precedenza, illustrando dettagliatamente i test effettuati in fase sperimentale.

Sebbene l'obiettivo finale riguardi la creazione di modelli che riconoscono le emozioni umane nel testo, il seguente lavoro si dedica anche allo sviluppo di attività descrittive e di analisi esplorative preliminari, in un'ottica di una completa esaustività del tema trattato.

Verrà innanzitutto descritto il dataset utilizzato per le analisi, poi si susseguiranno appositi paragrafi dedicati alle singole tecnologie e i relativi parameter tuning utili ad ottimizzare i modelli.

Infine, l'ultima parte è dedicata ad una valutazione sperimentale dei test eseguiti nonché una discussione dei risultati ottenuti.

5.1. IL DATASET

TwIT è il nome del dataset utilizzato in questo lavoro per condurre gli esperimenti di seguito descritti ed è un ampio set di dati composto da tweet in lingua italiana con la relativa emozione annotata manualmente.

Tale dataset è composto da 3108 records e 3 features:

- Id, numero progressivo associato a ciascun tweet
- Text, ossia il testo scritto dall'utente all'interno del tweet
- Emotion, l'emozione associata a ciascun tweet

Le emozioni presenti nel dataset sono le sei visibili in figura 11, associate manualmente ai singoli tweet mediante l'utilizzo di alcune particolari emoji come chiavi di ricerca.

Figura 11: Emotion and Emoji mapping

Emotion	Emojis
happiness	❤️ 😊
trust	👉
sadness	😓 😞
anger	😡 😠
fear	😱 🤯
disgust	🤢

Fonte: Potena D. et. al.,2022

Mediante le API di Twitter è stato possibile utilizzare le emoji sopra riportate per raccogliere diversi tweet appartenenti a una grande varietà di argomenti, per poi procedere ad una fase di elaborazione e pulizia del testo per rimuovere tweet inadatti e contenenti rumore [11].

Il risultato finale è un dataset in cui ad ogni tweet è associata una emozione, visibile in figura 12. La colonna 'Emotion' è stata trasformata da stringa a numerica per facilitare le operazioni future, partendo da zero per happiness, fino a cinque per l'emozione disgust.

Figura 12: TwIT Dataset

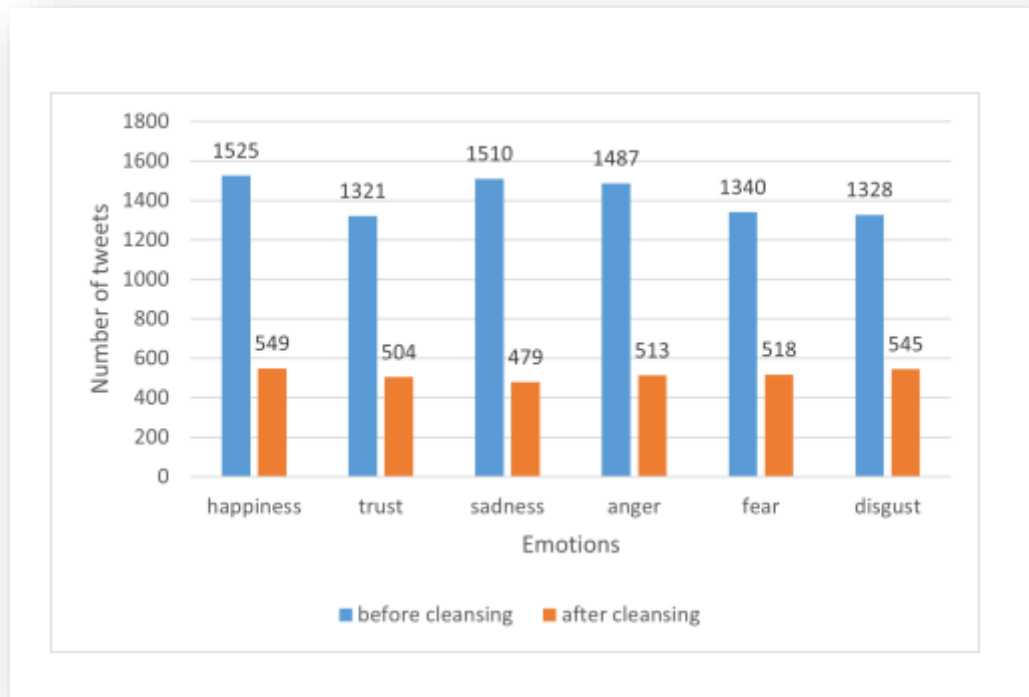
	Text	Emotion
Id		
2932	Hanno imparato bene dall'America. \r\nQuesta è...	5
2125	Io non ci voglio più venire a Roma	4
136	Il buongiorno che speravo di dare da molto tem...	0
2154	Subito Mary, non prendere mi raccomando	4
678	Già non mi perdevo una puntata prima figurati ...	1
1999	Ti odiamo proprio, altro che cattiva	3
610	Ciao, molto bello	1
561	GRANDE LISO \r\nGIU' LE MANI DALLE NOSTRE TERR...	1
2077	ah, ad esempio io ancora non sapevo che l'Asal...	4
399	Finalmente ho iliad	0

5.2. ANALISI DESCRITTIVA

Dopo aver importato il dataset descritto in precedenza, è stata condotta una prima fase di analisi esplorativa preliminare per valutare la coerenza e l'integrità dei dati rispetto agli obiettivi finali.

Le suddette analisi confermano l'assenza di valori nulli e inoltre un'analisi grafica della distribuzione di frequenza dei tweet in base alla relativa emozione è stata utile per capire se il dataset era bilanciato. In caso contrario, era necessario sottoporre i dati ad operazioni di rebalancing mediante oversampling delle classi minoritarie oppure undersampling delle classi maggioritarie.

Figura 12: Frequency distribution of tweets by emotion



Fonte: Potena D., et al., 2022

In figura 12 si nota come le colonne arancioni, che indicano il numero di tweets per emozione dopo la fase di cleaning, siano bilanciate tra loro, con happiness e disgust le più popolate e sadness la meno numerosa, ma comunque in linea con le altre.

Una volta appurato il bilanciamento del dataset nelle sei classi, è stata condotta una fase esplorativa/descrittiva volta ad illustrare la frequenza delle singole parole individuando quelle più utilizzate per esprimere una data emozione.

Tabella 1: le 10 parole più frequenti in Happiness

	HAPPINESS	TRUST	SADNESS	ANGER	FEAR	DISGUST
GRAZIA	68	24	13	8	3	3
BUONGIORNO	68	4	3	0	0	3
BELLO	66	11	5	1	2	1
FELICE	29	2	4	0	0	1
SABATO	29	0	0	0	1	0
AMORE	27	8	7	1	2	0
SETTIMANA	24	3	3	7	1	2
MERAVIGLIOSO	21	11	1	1	0	1
ABBRACCIARE	19	2	6	0	1	1
CUORE	16	5	9	2	1	1

Tabella 2: le 10 parole più frequenti in Trust

	HAPPINESS	TRUST	SADNESS	ANGER	FEAR	DISGUST
BRAVO	4	69	1	2	1	2
COMPLIMENTO	3	46	0	0	2	2
GRANDE	12	42	6	3	7	2
BRAVISSIMA	2	23	0	0	0	0
OTTIMO	0	16	0	3	1	0
CONGRATULAZIONE	2	16	1	0	0	1
BELLISSIMO	6	15	0	0	0	0
FANTASTICO	6	14	1	2	0	0
MERITARE	2	10	0	0	1	0
PAZZESCO	6	10	2	3	0	4

Tabella 3: le 10 parole più frequenti in Sadness

	HAPPINESS	TRUST	SADNESS	ANGER	FEAR	DISGUST
DISPIACERE	0	2	29	1	7	6
PIANGERE	1	0	22	0	6	1
POVERO	0	0	19	11	7	1
TRISTE	0	0	18	0	0	0
MANCO	0	0	17	2	4	2
SPERARE	6	4	17	5	12	8
SENZA	6	1	16	11	18	12
MALE	0	0	15	9	17	7
TRISTEZZA	0	0	12	0	3	1
SCUSA	2	1	11	3	0	2

Tabella 4. Le 10 parole più frequenti in Anger

	HAPPINESS	TRUST	SADNESS	ANGER	FEAR	DISGUST
VERGOGNA	0	0	0	31	4	13
VACCINARE	0	1	3	24	10	18
BASTARE	0	1	7	23	10	13
NORIMBERGA	0	0	0	17	0	1
SOLDI	1	0	0	15	0	4
ITALIA	4	9	7	14	10	6
PAESE	2	3	6	11	10	12
CRIMINALE	0	0	0	10	0	1
TASSE	0	0	0	10	0	0
GALERA	0	0	0	9	0	3

Tabella 5: le 10 parole più frequenti in Fear

	HAPPINESS	TRUST	SADNESS	ANGER	FEAR	DISGUST
SOLO	6	14	24	27	22	39
PAURA	0	3	2	3	18	0
MALE	0	0	15	9	17	7
MONDO	4	7	7	6	12	6
ANSIA	0	0	5	0	10	1
AMICO	10	4	4	4	9	2
MORIRE	4	0	7	7	8	4
ODDIO	4	0	0	1	8	4
PROBLEMA	3	0	4	8	7	2
PANDEMIA	0	1	1	0	6	2

Tabella 6: le 10 parole più frequenti in Disgust

	HAPPINESS	TRUST	SADNESS	ANGER	FEAR	DISGUST
SCHIFARE	0	0	4	9	2	64
VOMITARE	0	0	1	0	0	26
PURO	3	5	4	8	5	20
POLITICA	0	0	1	4	1	11
SCHIFOSO	0	0	1	1	2	10
PEGGIO	0	2	0	5	5	10
VOTARE	0	2	0	8	2	8
FALSO	0	0	0	0	0	7
MANGIARE	2	0	1	3	1	9
PROPRIO	4	12	9	11	16	20

Le tabelle propongono una distribuzione di frequenza per le prime dieci parole maggiormente utilizzate in ogni emozione, nonché la loro ricorrenza anche nelle altre emozioni così da evidenziare maggiormente la relazione tra la parola e la classe in cui la frequenza è maggioritaria.

È possibile osservare che in tabella 1 le parole ‘Buongiorno’, ‘Bello’, ‘Felice’, ‘Meraviglioso’ ed altre, ricorrono molte volte nei tweet etichettati come happiness, mentre il loro utilizzo si riduce notevolmente nelle classi anger, fear e disgust.

I medesimi risultati si ottengono anche con le parole più frequenti nelle altre emozioni, con il loro utilizzo che si riduce notevolmente in contesti differenti, fino al caso estremo in cui la parola non risulta nemmeno una volta utilizzata.

Al termine della fase esplorativa, è possibile confermare che esistono alcune parole ampiamente utilizzate in tweet con etichette simili, il cui utilizzo si riduce notevolmente in testi contenenti altre emozioni.

5.3. UNSUPERVISED LEARNING

Dopo un'accurata analisi preliminare, si procede all'implementazione di tecniche di apprendimento non supervisionato con l'obiettivo di verificare se esiste una relazione tra vicinanza semantica e vicinanza emotiva delle parole presenti nei tweet.

Per constatare ciò, l'algoritmo Word2Vec ha permesso di associare le parole a vettori multidimensionali e poi, tecniche di clustering hanno reso possibile il raggruppamento delle stesse. Infine, un'apposita tabella pivot sintetizza i risultati ottenuti dalla quale è possibile trarre conclusioni ragionate rispetto all'obiettivo iniziale.

Nei paragrafi successivi si illustra il procedimento seguito e si descrivono le variazioni dei parametri più significativi in grado di ottimizzare i risultati e scovare relazioni nascoste nei dati di partenza.

5.3.1. Word2Vec

Sebbene Word2Vec permetta di ottenere rappresentazioni vettoriali multidimensionali per ogni singola parola, è necessario prima creare un corpus di testo ad hoc, dove ogni singolo elemento è nel formato idoneo alla sua successiva elaborazione algoritmica.

Viene così avviata la fase di word embedding pre-processing, così composta:

- rimozione di caratteri speciali e singoli caratteri
- Sostituzione degli spazi multipli con spazi singoli
- Conversione in caratteri minuscoli di tutti gli elementi
- Rimozione di numeri da stringhe
- Rimozione di stop words
- Lemmatizzazione
- Tokenizzazione

Al termine del processo si ottiene un corpus di testo composto da tanti token quanti sono le diverse parole presenti nei tweet e, per ogni token, viene creato il suo corrispettivo vettore numerico.

Word2Vec restituisce un modello contenente tutti i word embedding associati alle parole del testo e, a seconda della dimensione vettoriale, essi si distribuiranno in maniera differente nello spazio.

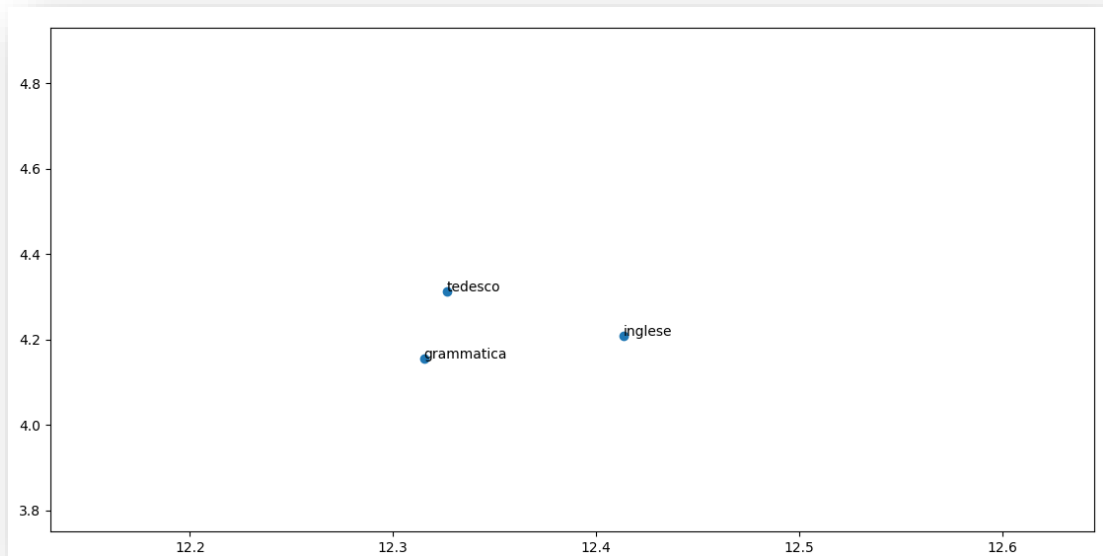
La qualità del modello è però sensibile a parametrizzazione e per trovare quello che definisce meglio il contesto semantico delle parole è stato necessario testare l'algoritmo cambiando ripetutamente alcuni parametri significativi.

Nel dettaglio, visti i numerosi esperimenti condotti in questa sede, è possibile dire con certezza che il parametro che più influenza i risultati finali è la dimensione associata a ciascun vettore. Tra i tanti test condotti, si è notato che un numero compreso tra 100 e 300 sia la giusta dimensione da attribuire ai vettori per ottenere distribuzioni spaziali sensate. Aumentando tale valore al di sopra di 300 dimensioni si otterranno vettori troppo sparsi che non riusciranno a cogliere le relazioni presenti nei dati. Riducendo invece la dimensione a valori troppo piccoli, si otterranno tanti vettori vicini e sarà difficile distinguere quali saranno correlati e quali no.

Al fine di una fase sperimentale dettagliata, si è optato per un modello composto da vettori distribuiti in 128 dimensioni, calibrando poi i restanti parametri per ottenere la miglior rappresentazione degli stessi. In particolare, con 128 come valore di vector size, si è fissato il numero del parametro window a 5 utilizzando come architettura log-lineare Continuous Bag Of Word.

Sebbene non esistano metriche specifiche per verificare i risultati dell'algoritmo, mediante alcune visualizzazioni grafiche è stato possibile determinare con buona approssimazione il successo o meno del modello. In figura 13 si riporta una zona dello spazio vettoriale creato in cui è possibile osservare parole che sono semanticamente simili tra loro.

Figura 13: Word Embedding visualization



Essendo le dimensioni dei vettori superiori a due dimensioni, per rappresentarli graficamente si è optato per la tecnica t-SNE, t-distributed Stochastic Neighbor Embedding, ossia un algoritmo di riduzione della dimensionalità simile alla PCA ma che si presta particolarmente bene nel contesto degli embedding ad alte dimensioni cercando di preservarne la struttura locale.

Altre configurazioni di word embedding mediante Word2Vec sono state testate e sono risultate altrettanto efficaci. In particolare, aumentando la dimensione dei vettori a 250 e utilizzando l'architettura log-lineare opposta alla precedente, cioè Skip-gram con 10 parole di contesto nel parametro window, si raggiungono buoni risultati. In questo caso però, si sono testati i valori finali mediante approcci di similarità e dissimilarità. Data cioè una parola in input e la sua rappresentazione vettoriale nello spazio, la funzione *most similar* permette di ottenere in output le parole più simili (in termini di vicinanza spazio-vettoriale) a quella di input. Il test è stato eseguito sulla parola 'amore' e le tre parole più vicine ad essa sono state 'amo', 'felicità' e 'stupendo'.

Al contrario, la funzione *doesn't match* prende in input una lista di parole e restituisce in output quella più distante in termini spazio-vettoriali. Il test in particolare è stato fatto con

le parole ‘amore’, ‘felicità’ e ‘ospedale’ e la funzione ha restituito in output ‘ospedale’, considerata la più lontana e quella semanticamente più discordante.

Tale analisi di similarità/dissimilarità è stata riproposta anche al modello precedente con 128 dimensioni e i risultati sono pressoché simili. Si è quindi optato per mantenere il primo come modello principale di riferimento.

Accertati quindi gli ottimi risultati del modello Word2Vec a 128 dimensioni, le distribuzioni vettoriali delle parole sono state salvate in un file csv e successivamente utilizzate come input per algoritmi di clustering.

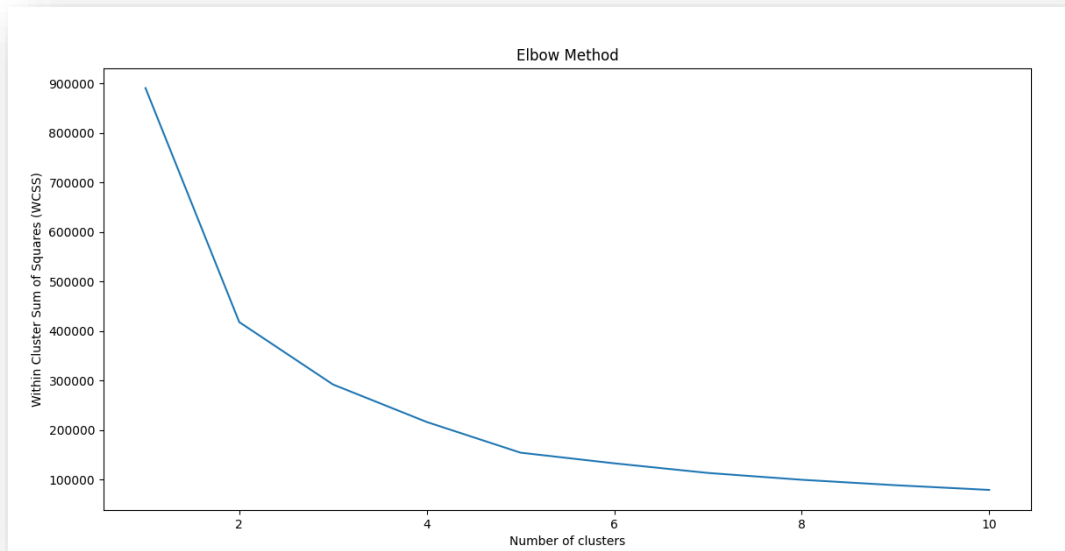
5.3.2. Clustering

In questa sede, l’obiettivo della cluster analysis è quello di raggruppare i vettori di parole ottenuti nella fase precedente e studiare gli elementi interni a ciascun cluster, così da capire se esiste una corrispondenza tra vicinanza semantica e vicinanza emotiva.

Tra le varie tecniche di clustering, si è optato per k-means come algoritmo principale per creare i cluster di parole e, sebbene le etichette che identificano le emozioni suggeriscano sei cluster, è stato comunque implementato un elbow method per verificare se sei è un buon numero per l’iper-parametro k.

Dal grafico presente in figura 14 si nota come la curva tende ad appiattirsi da quattro cluster in poi e non avendo un numero ottimale, si è deciso di testare diversi valori di k, in particolare quattro, cinque e sei.

Figura 14: Determine optimal cluster number with elbow method



Inoltre, si è testato l'algoritmo passandogli sia il vettore originale da 128 dimensioni che preserva tutta l'informazione, sia vettori ridotti a due o tre dimensioni. Tali approcci si differenziano in quanto nel primo caso, si potrebbe andare incontro ad un problema di course of dimensionality mentre nel secondo caso si potrebbe perdere tanta informazione.

Dopo numerosi test effettuati, si può confermare che i vettori originali non garantiscono buoni risultati nella creazione dei gruppi, con cluster molto popolati ed alcuni invece composti da sole poche unità. Anche riducendo o aumentando il numero di cluster rispetto a quelli suggeriti dall'elbow method non si raggiungono raggruppamenti soddisfacenti.

I risultati precedenti suggeriscono quindi di testare l'algoritmo riducendo, anche gradualmente, il numero di dimensioni dello spazio vettoriale per ottenere una configurazione finale più omogenea, con cluster più definiti e facilmente interpretabili.

Si è quindi optato per testare sia valori intermedi, sia valori molto bassi, implementando anche altre tecniche di riduzione della dimensionalità oltre a t-SNE, quale PCA e UMAP.

Si è riscontrato un notevole miglioramento dei cluster riducendo i vettori a dimensioni molto basse, mentre per quanto riguarda dimensioni intermedie quali 50 e 30 i risultati sono pressoché simili a quelli con l'utilizzo dei vettori originali.

In particolare, osservando anche il valore della silhouette score, che è una delle metriche utilizzate in questa sede per valutare l'analisi cluster, si può confermare che il miglior risultato è stato ottenuto fissando sei come numero di cluster, in uno spazio vettoriale ridotto a tre dimensioni mediante t-SNE. In questo caso specifico, la silhouette score è risultata 0.48, il numero più alto che questa metrica ha registrato rispetto alle configurazioni testate in precedenza.

Si è quindi scelto di concludere tale analisi con la configurazione che ha offerto sia risultati più solidi dal punto di vista della silhouette score, sia dalla formazione finale dei cluster, che sono risultati quantitativamente più omogenei.

Ne consegue quindi una analisi incrociata dei dati ottenuti dalla cluster analysis, proponendo una dashboard dinamica in grado di mettere in evidenza quali parole sono contenute in ciascun cluster, verificando se quest'ultimi possono ricondursi a raggruppamenti emotivi.

Tabella 7: pivot table of word distribution in clusters

Conteggio di HAPPINESS		Emozioni							
Cluster=6--- TSNE 3 dimensioni	ambiguo	Anger	Disgust	Fear	Happiness	Sadness	Trust	Totale complessivo	
0		13	12	2		3	1	31	
1		19	1	4	1		2	28	
2		12	4	2	3		1	26	
3		202	3	4	2	43	4	278	
4		155	25	18	5	10	10	234	
5		135	21	12	9	4	8	197	
Totale complessivo		536	66	42	20	60	25	794	

La tabella 7 propone un riassunto di quante e quali parole sono contenute in ciascun cluster, etichettandole in una delle sei emozioni se la sua frequenza relativa superava una certa soglia, altrimenti classificata come ambigua in quanto presente in modo frequente in più emozioni.

In questo caso, se i cluster ottenuti contengono esattamente parole con frequenze relative riconducibili ad una emozione specifica, si può constatare che in quel determinato cluster

emerge una correlazione tra vicinanza semantica e vicinanza emotiva. Se, al contrario, i cluster sono eterogenei, ovvero composti da molteplici parole ambigue o con frequenze relative discordanti, si può confermare che non esiste alcun tipo di relazione tra la semantica e la relativa emozione associata.

Una volta prese le parole più frequenti per ogni emozione, che si ricorda essere un'informazione già acquisita in fase di analisi descrittiva, si sono calcolate le frequenze assolute e relative, fissando una soglia variabile come livello al di sopra del quale la parola poteva essere considerata appartenente ad un'emozione. Tale soglia è stata fatta variare da un minimo di 0.3 fino ad un massimo di 0.8 per osservare come cambiavano i risultati finali.

In tabella 7 è stata utilizzata una soglia pari a 0.6 in quanto garantiva il minor numero di ambigui, anche se pur sempre un numero molto elevato. Si osserva come tre cluster contengono gran parte delle parole presenti, mentre i restanti tre risultano poveri di unità. Le parole appartenenti alle emozioni happiness e trust si distribuiscono principalmente nel cluster tre, mentre le parole appartenenti alle emozioni anger, disgust e sadness si distribuiscono perlopiù nei cluster quattro e cinque. Non si riscontra invece una equa distribuzione per l'emozione fear, le cui parole risultano dislocate equamente in più clusters.

Aumentando la soglia a 0.7 o più, si riscontra un numero altamente elevato di parole ambigue la cui frequenza relativa non raggiunge tale soglia in nessuna delle emozioni presenti creando situazioni di incertezza. Al contrario, riducendo la soglia a valori più piccoli come 0.4 o 0.3, esistono numerose parole che superano tale soglia in più emozioni e quindi risulta difficile classificarle in una piuttosto che un'altra.

5.4. SUPERVISED LEARNING

Le tecniche supervisionate permettono di raggiungere l'obiettivo principale del progetto, ossia la creazione di due differenti modelli di classificazione in grado di riconoscere le emozioni nel testo.

Il primo classificatore è stato sviluppato con GPT-3, uno dei modelli linguistici più all'avanguardia in ambito NLP, in grado di raggiungere risultati straordinari in differenti attività che comprendono l'elaborazione del linguaggio.

Il secondo invece si è appoggiato a BERT, un altro grande modello linguistico proposto da Google che garantisce ottime performance anche se presenta un'architettura più contenuta rispetto GPT-3.

5.4.1. GPT-3

La realizzazione di modelli di classificazione mediante GPT-3 risulta di facile implementazione in quanto, non essendo un framework open source, non è possibile cambiare a proprio piacimento i parametri. Ne deriva che l'unico parametro configurabile risulta essere il motore di elaborazione dei dati, dal più piccolo ma più economico, fino al più grande ma il più prestante. Gli esperimenti effettuati in questa sede hanno utilizzato due motori resi disponibili da OpenAI, Curie e Davinci. Essendo però Davinci quello con più parametri che fornisce prestazioni migliori, si riportano per semplicità solo i risultati ottenuti dal modello addestrato con quest'ultimo motore.

L'approccio seguito per la creazione del modello prevede la suddivisione del set di training per l'80% dei dati, mentre il restante 20% forma il test set sulla quale poi valutarne le performance.

Mediante la chiave API fornita da OpenAI è stato possibile fare Fine-Tuning sui dati di addestramento, ottenendo indietro un ID univoco associato al nostro particolare modello di classificazione.

Infine, il test set è stato utile per fare le previsioni su dati nuovi che il modello non aveva mai visto, calcolando di conseguenza le metriche più comuni per valutarne le performance, quale accuracy, precision, recall e F1-score.

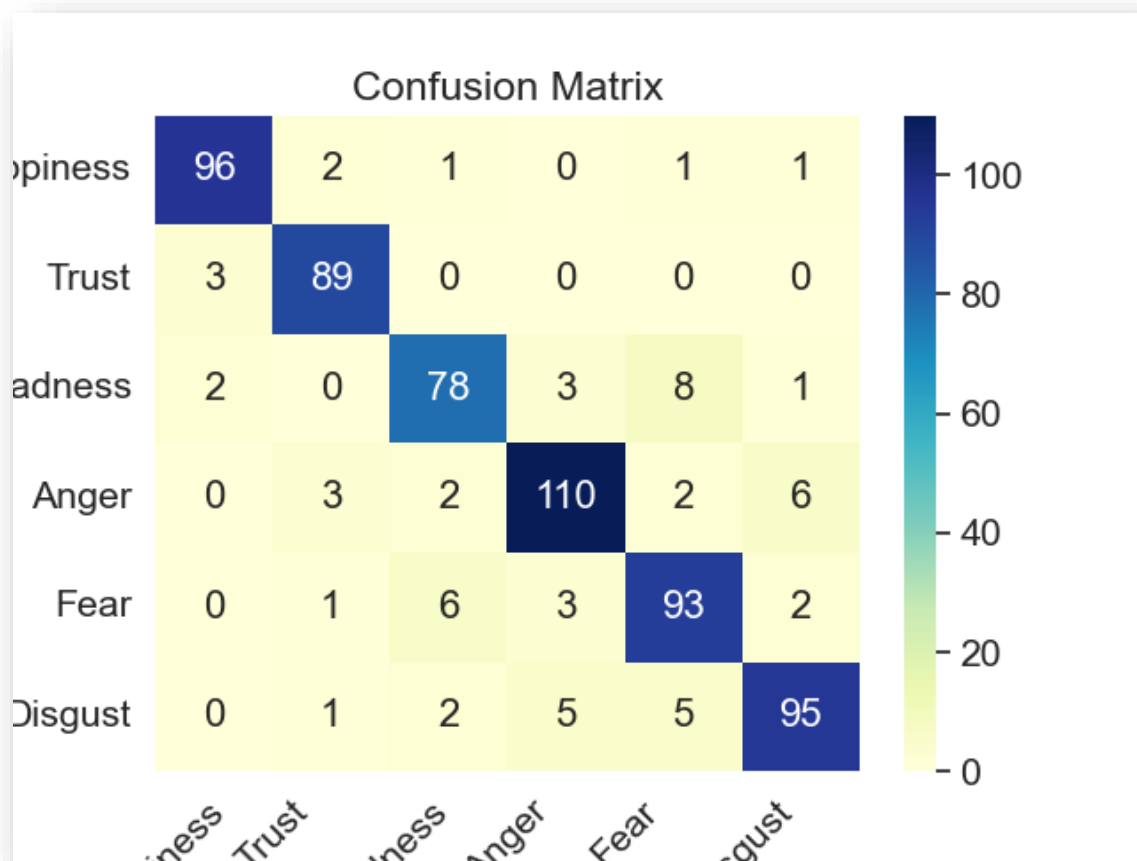
Tabella 8: GPT-3 classification report

	precision	recall	f1-score	support
0	0.94	0.96	0.95	117
1	0.97	0.93	0.95	90
2	0.86	0.90	0.88	92
3	0.86	0.88	0.87	104
4	0.92	0.82	0.87	114
5	0.87	0.93	0.90	101
accuracy			0.90	618
macro avg	0.90	0.90	0.90	618
weighted avg	0.90	0.90	0.90	618

In tabella 8 è riportato il classification report del modello creato, che, come si può notare, raggiunge un livello di accuratezza molto elevato classificando correttamente il 90% dei dati di test. Anche l’F1-score, che tiene in considerazione precision e recall dei dati di test per singola etichetta, offre risultati soddisfacenti, sfiorando il 95% in happiness e trust.

Se nel dettaglio si vuole osservare quali dati sono stati classificati erroneamente rispetto la classe vera, è possibile studiare la matrice di confusione riportata in figura 15.

Figura 15: GPT-3 confusion matrix



La confusion matrix fa notare come le emozioni sadness e anger siano le più difficili da riconoscere, sbagliando rispettivamente quattordici e tredici osservazioni, mentre solo tre unità sono mal classificate in trust, raggiungendo una precision molto alta, pari al 97%.

Questi risultati verranno poi confrontati con il successivo modello creato con Google BERT, che offre una maggior flessibilità nel definire i valori dei singoli parametri.

5.4.2. Google BERT

Il modello di classificazione creato con BERT è altrettanto valido in quanto permette all'utente maggiore flessibilità nell'attività di parameter tuning e garantisce un'ottimizzazione delle performance del classificatore.

Dopo aver ripreso gli stessi dati di training e test set utilizzati in GPT-3, è stato caricato il modello pre-addestrato 'bert base multilingual cased' in grado di riconoscere anche la lingua italiana nell'attività di comprensione del linguaggio.

Successivamente, strati densi aggiuntivi sono stati creati al di sopra del modello base mediante il framework Tensorflow e Keras come libreria open source. In particolare, diverse configurazioni sono state testate, ma in questo caso solo una di esse ha dato buoni risultati mentre le altre cadevano in problemi di overfitting del modello.

La miglior configurazione prevede due strati densi aggiuntivi con rispettivamente 64 e 6 neuroni per strato, utilizzando pure un dropout pari a 0.2. In sintesi, dropout o strato di abbandono, aiuta a prevenire l'overfitting del modello ignorando alcuni neuroni durante la fase di addestramento, cioè i loro pesi non vengono propagati al livello successivo [35].

La fase di parameter tuning è risultata assai complessa per via anche di alcuni valori di max length, epochs e learning rate. Max length, il parametro che definisce la lunghezza massima ammissibile della frase oltre al quale essa viene troncata, è risultato ottimale quando fissato a 64. Anche il numero di epoche è risultato fondamentale per ottenere performance di livello, testando diversi valori da due fino a dieci. Oltre a dieci, si presumeva che il modello potesse overfittare i dati e quindi non sono stati fatti ulteriori test.

Il training è stato effettuato utilizzando Adam, un algoritmo di ottimizzazione alternativo alla discesa del gradiente stocastico in grado di aggiornare i pesi della rete in base ai dati di addestramento che gli vengono forniti. Seguendo il consiglio dei ricercatori Google, si è deciso di far variare il learning rate in un range di tre valori, $1e-5$, $2e-5$ e $3e-5$ con risultati pressoché simili tra loro.

Tabella 9: BERT classification report

	precision	recall	f1-score	support
Happiness	1.00	0.93	0.96	117
Trust	0.93	0.97	0.95	90
Sadness	0.92	0.86	0.89	92
Anger	0.75	0.90	0.82	104
Fear	0.92	0.80	0.85	114
Disgust	0.86	0.92	0.89	101
accuracy			0.89	618
macro avg	0.90	0.90	0.89	618
weighted avg	0.90	0.89	0.90	618

Impostando il modello come detto in precedenza, addestrandolo sul training e testandolo sul test set, si ottengono risultati soddisfacenti in termini di accuratezza e altre metriche riportate nel classification report in tabella 9.

L'accuratezza è del 89% e anche in questo caso le emozioni più riconoscibili sono happiness e trust, con rispettivamente 96% e 95% come valore dell'F1-score.

Anger rimane l'emozione più ostica per entrambi i modelli, in questo caso con un valore F1 pari a 82%, il più basso rispetto l'F1 delle altre emozioni.

Modificando l'architettura della rete, ossia aumentando o diminuendo il numero di strati densi o semplicemente aumentando il numero di neuroni presenti nei singoli strati, il

modello produce overfitting dei dati di training e genera livelli di accuratezza molto bassi nei nuovi dati di test.

Ciò è stato confermato, sia aumentando l'architettura di uno strato denso aggiuntivo, sia aumentando il numero di neuroni a 128 per lo strato intermedio e con o senza dropout il risultato non è comunque buono.

Si è deciso di confrontare il modello GPT-3 con l'unica configurazione di BERT che offre performance accettabili. Il prossimo paragrafo prevederà una discussione dei risultati ottenuti dai due modelli, con particolare riferimento ai costi sostenuti per il loro sviluppo.

5.5. VALUTAZIONE DEI RISULTATI OTTENUTI

Entrambi i modelli proposti garantiscono ottimi risultati in termini di accuratezza e, sebbene esista una differenza percentuale minima dell'1% tra i due classificatori, è difficile stabilire quale dei due possa essere il migliore.

Non esiste un modello che prevale nettamente sull'altro, ma dipende dal contesto di riferimento e dall'obiettivo di business che il management si propone di raggiungere. Certo è che entrambi funzionano molto bene nel loro intento e possono essere uno strumento assai importante per qualsiasi realtà aziendale data-oriented, che fa cioè del dato un elemento imprescindibile per il raggiungimento di vantaggi competitivi sostenibili nel tempo.

I modelli pre-addestrati utilizzati si differenziano dal punto di vista pratico perché possiedono costi economici e computazionali differenti, nonché tempi di training che differiscono di vari ordini di grandezza.

Nello specifico, il progetto in questione ha evidenziato alcune discrepanze tra le due implementazioni, attribuendo alcuni vantaggi e svantaggi ad una tecnologia piuttosto che all'altra.

GPT-3 processa i dati di training mediante strumenti cloud senza lavorare in locale, garantendo un costo computazionale estremamente basso. Anche il learning-time è conveniente in quanto riesce a creare un modello da zero in poco tempo, a seconda della quantità di dati passati durante la fase di addestramento. In questo caso specifico il modello è stato creato in circa 20 minuti.

D'altro canto, lo svantaggio più grande è che il suo utilizzo è dettato da condizioni economiche che, peraltro, variano a seconda del motore scelto in fase di addestramento. Non solo il fine tuning è a pagamento, ma anche la successiva fase di predizione ha un costo unitario variabile per singola frase, in base al numero di token che essa stessa possiede. In questo lavoro il costo è stato facilmente sostenibile senza particolari problemi, ma in ottica aziendale durante lo sviluppo di modelli con milioni e milioni di dati di training è necessario considerare questo aspetto assai caratterizzante di GPT-3.

Google BERT invece possiede caratteristiche opposte a GPT-3 essendo fornito di librerie open source che permettono la sua implementazione. I vantaggi di questo approccio sono senza dubbio interessanti, in quanto permettono di definire architetture personalizzate a specifici casi d'uso, ma richiedono notevoli competenze tecniche nella fase di parameter tuning, altrimenti si potrebbero creare modelli poco performanti.

A differenza del modello precedentemente creato con GPT-3, quello messo a punto con Google BERT ha impiegato molto più tempo per la fase di addestramento, calcolando un learning time di circa 2 ore per epoca e, considerando che le epoche erano 10, l'algoritmo ha impiegato circa un giorno per completare il ciclo.

Tale tempistica può essere notevolmente ridotta se si fruttano le capacità computazionali della Graphics Process Unit, cioè la scheda grafica in grado di processare più attività in parallelo riducendo i tempi di lavoro.

Nonostante ciò, a differenza di GPT-3 l'addestramento avviene in locale sfruttando le componenti hardware del computer e questo è sicuramente svantaggioso perché, oltre a spendere potenza computazionale propria, implica l'acquisto di componenti costosi se si vogliono mettere a punto modelli molto grandi.

Maggiore è la scala, maggiori sono le performance dei modelli ma, al contempo, maggiori sono i costi economici per l'acquisizione di infrastrutture adeguate a sostenerne lo

sviluppo. Risulta quindi di fondamentale importanza considerare la possibilità di dover sostenere un investimento iniziale corposo se si opta per BERT o altri modelli open source.

6. CONCLUSIONI

L'obiettivo del lavoro era quello di creare modelli di classificazione in grado di riconoscere le emozioni umane nei testi in lingua italiana e, alla luce dei risultati ottenuti, si può confermare che tale obiettivo è stato raggiunto.

Dopo una parte introduttiva in cui è stato definito il contesto di riferimento e le sfide che le imprese si trovano a dover affrontare, sono stati trattati i modelli teorici più promettenti in ambito emozionale. A partire dagli studi di Paul Ekman, è risultato di fondamentale importanza approfondire la ruota delle emozioni di Robert Plutchik che ha permesso di comprendere a fondo come queste venivano categorizzate. I modelli emozionali hanno fornito un sostegno concreto allo studio delle emozioni e al loro manifestarsi, illustrando dal punto di vista teorico quali relazioni sono insite in ciascuna di essi.

L'analisi delle emozioni è stata effettuata attraverso lo sviluppo di modelli predittivi basati su algoritmi di deep learning allo stato dell'arte. Preliminarmente è stato analizzato il dataset utilizzato per gli esperimenti, composto da 3108 tweet in lingua italiana e la relativa emozione associata. È stata poi condotta una fase preliminare di analisi descrittiva ed esplorativa con l'intento di scovare pattern e relazioni nascoste nei dati. In particolare, l'analisi era finalizzata alla creazione di tabelle rappresentanti le distribuzioni di frequenza delle prime dieci parole maggiormente utilizzate in ogni emozione e la loro ricorrenza anche nelle altre emozioni. Al termine della fase esplorativa, è stato possibile confermare l'esistenza di alcune parole ampiamente utilizzate in tweet con etichette simili, il cui utilizzo si è ridotto notevolmente in testi contenenti altre emozioni. Il dataset è quindi stato analizzato mediante tecniche non supervisionate e supervisionate.

Per la parte non supervisionata, l'obiettivo è stato quello di verificare se esisteva una relazione tra vicinanza semantica e vicinanza emotiva delle parole presenti nei tweet. Dopo diverse implementazioni di Word2Vec è stato appurato che quest'ultimo performava al meglio trasformando le parole in vettori composti da 128 dimensioni. Fra le tecniche di clustering testate, k-means è risultato il miglior algoritmo per raggruppare i word embedding.

Per la parte supervisionata, l'obiettivo è stato quello di creare modelli di classificazione delle emozioni per testi in lingua italiana. In particolare, sono stati creati due classificatori utilizzando modelli linguistici differenti. Il primo con GPT-3 ha raggiunto un'accuracy pari al 90% nei dati di test. Il secondo con Google BERT ha invece raggiunto un livello di accuracy del'89%, un punto percentuale in meno rispetto GPT-3. Entrambi comunque sono stati soddisfacenti nel riconoscere l'emozione che governava il testo, confermando il raggiungimento dell'obiettivo iniziale.

Infine, i due modelli sono stati comparati dal punto di vista di tre indicatori: tempi di addestramento, costi economici e costi computazionali.

In sintesi, il modello sviluppato con GPT-3 è stato più efficiente dal punto di vista computazionale, ma non dal punto di vista economico in quanto ha sottratto più risorse rispetto l'altro modello. Invece per quanto riguarda i tempi di training, è stato nettamente più performante rispetto BERT.

D'altro canto, Google BERT è stato poco efficiente sia dal punto di vista computazionale che per quanto riguarda i tempi di training, impiegando notevoli quantità di tempo per l'addestramento del modello. I costi economici sono stati pressoché nulli anche se è bene considerare la possibilità di dover sostenere un investimento iniziale per l'acquisto di infrastrutture di base se si opta per questo modello o altri modelli open-source.

6.1. Sviluppi futuri

Il seguente paragrafo propone alcune riflessioni circa gli sviluppi che in futuro si potrebbero apportare al seguente lavoro, migliorandolo mediante nuove tecnologie e/o modelli linguistici, oppure espandendolo definendo nuovi obiettivi di analisi.

Quindi, possibili sviluppi futuri del seguente lavoro includono:

- Test dei classificatori su altri dataset che non necessariamente comprendono tweet, ma anche post su Facebook, o siti di recensioni come TripAdvisor;
- Creare un modello di classificazione utilizzando l'evoluzione di GPT-3, ossia GPT-4 in uscita nel secondo semestre del 2023;

- Implementazione di un sistema preliminare finalizzato a discriminare i messaggi oggettivi/neutri, che non presentano opinioni personali ed emozioni ma che esprimono solamente fatti, da messaggi soggettivi che realmente contengono emozioni;
- Addestramento di modelli di classificazione in grado di riconoscere anche più di una emozione nel testo, in un contesto multi-etichetta;

In conclusione, ci si pone l'obiettivo di estendere il seguente progetto, implementando non solo le tecniche più all'avanguardia ma modificando anche gli obiettivi iniziali di analisi, nell'ottica di un più completo ed esaustivo scopo di ricerca.

SITOGRAFIA & BIBLIOGRAFIA

[1] Talkwalker.com. Social Media Trends 2023 download. Talkwalker. Retrieved November 16, 2022, from <https://www.talkwalker.com/social-media-trends-download>

[2] Fogliotiziana, By, Ceciliascalzo, Ivananotarangelo, Giannantonio, M. D., & AlessandroDiPiero. (2021, October 4). Infografica: Quello Che Accade in 1 Minuto su internet nel 2021. Syrus. Retrieved November 24, 2022, from <https://syrus.blog/infografica-quello-che-accade-in-1-minuto-su-internet-nel-2021-17445.html>

[3] Redazione, di. (n.d.). *Cognitivismo.com*. Cognitivismo. Retrieved January 5, 2023, from <https://cognitivismo.com/2012/08/01/un-modello-dimensionale-delle-emozioni-integrazione-tra-le-neuroscienze-dellaffettivita-lo-sviluppo-cognitivo-e-la-psicopatologia/>

[4] Wikimedia Foundation. (2022, December 1). Paul Ekman. Wikipedia. Retrieved December 3, 2022, from https://it.wikipedia.org/wiki/Paul_Ekman

[5] Fiore, F. (2018, June 14). Paul Ekman: Biografia E Contributi Scientifici sullo studio delle Emozioni. State of Mind. Retrieved December 3, 2022, from <https://www.stateofmind.it/2018/06/paul-ekman-emozioni-espressioni-facciali/>

[6] Emozione nell'Enciclopedia treccani. nell'Enciclopedia Treccani. (n.d.). Retrieved December 2, 2022, from <https://www.treccani.it/enciclopedia/emozione/>

[7] La Ruota delle Emozioni di Plutchik. Mariangela De Rogatis Psicologa. (2021, September 6). Retrieved December 2, 2022, from <http://www.mariangeladerogatispsicologa.it/la-ruota-delle-emozioni-di-plutchik/>

- [8] Mordenti, A. (2022, September 15). Analisi Dati, tecniche E metodologie di applicazione. Big Data 4Innovation. Retrieved December 3, 2022, from <https://www.bigdata4innovation.it/big-data/analisi-dei-dati-tecniche-e-metodologie-per-la-sua-applicazione/>
- [9] Wikimedia Foundation. (2022, December 4). Twitter. Wikipedia. Retrieved December 5, 2022, from <https://it.wikipedia.org/wiki/Twitter>
- [10] Randazzo, R. M. (2022, October 5). Emotion recognition: Dalle Origini Ad Oggi. Blog e Risorse - Adv Media Lab. Retrieved December 5, 2022, from <https://blog.advmedialab.com/emotion-recognition-dalle-origini-ad-oggi>
- [11] Chiorrini A., Diamantini C., Mircoli A., Potena D., Storti E. (2022). “EmotionAlBERTo: Emotion Recognition of Italian Social Media Texts Through BERT”.
- [12] Intelligenza artificiale (IA): Cos'è, come funziona e come applicarla. Intelligenza Artificiale (IA): cos'è, come funziona e come applicarla. (n.d.). Retrieved December 6, 2022, from https://blog.osservatori.net/it_it/intelligenza-artificiale-funzionamento-applicazioni#significato
- [13] Boldrini, N., Stefano, Ramona, & Atti, G. (2022, November 4). Intelligenza artificiale - cos'è, come funziona, esempi 2022 e AI in Italia. AI4Business. Retrieved December 6, 2022, from <https://www.ai4business.it/intelligenza-artificiale/intelligenza-artificiale-cose/>
- [14] Machine learning. Machine Learning - an overview | ScienceDirect Topics. (n.d.). Retrieved December 6, 2022, from <https://www.sciencedirect.com/topics/psychology/machine-learning#:~:text=In%2C%201959%20Arthur%20Samuel%20defined,explicitly%20programmed%E2%80%9D%20%5B2%5D>
- [15] Redazione, Rinaldi, E., & Bilo, L. (2022, November 4). Machine learning: Cos'è e applicazioni dell'apprendimento Automatico. AI4Business. Retrieved December 7, 2022, from <https://www.ai4business.it/intelligenza-artificiale/machine-learning/machine-learning-cosa-e-applicazioni/>

[16] Mitchell M. T. (1997, March 1).” Machine Learning: A multistrategy approach”

[17] Castigli, M. (2022, May 26). Deep learning: Cos'è l'apprendimento Profondo. Big Data 4Innovation. Retrieved December 7, 2022, from <https://www.bigdata4innovation.it/intelligenza-artificiale/deep-learning-cose-ed-esempi-dellapprendimento-profondo%E2%80%A8/>

[18] Wikimedia Foundation. (2022, November 3). Rete Neurale Artificiale. Wikipedia. Retrieved December 7, 2022, from https://it.wikipedia.org/wiki/Rete_neurale_artificiale

[19] Ognjanovski, G. (2020, June 7). Everything you need to know about neural networks and backpropagation-machine learning made easy... Medium. Retrieved December 7, 2022, from <https://towardsdatascience.com/everything-you-need-to-know-about-neural-networks-and-backpropagation-machine-learning-made-easy-e5285bc2be3a>

[20] Yse, D. L. (2019, April 30). Your guide to natural language processing (NLP). Medium. Retrieved December 8, 2022, from <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1>

[21] Altobello, G., & engineer, D. e A. I. (2022, August 22). Natural language processing, Cos'è, come funziona, applicazioni e python. AI4Business. Retrieved December 8, 2022, from https://www.ai4business.it/intelligenza-artificiale/natural-language-processing-tutto-quello-che-ce-da-sapere/#Cos8217e_il_Natural_Language_Processing

[22] Brownlee, J. (2019, August 7). What are word embeddings for text? MachineLearningMastery.com. Retrieved December 8, 2022, from <https://machinelearningmastery.com/what-are-word-embeddings/>

[23] Wikimedia Foundation. (2022, November 22). Transformer (Machine Learning Model). Wikipedia. Retrieved December 9, 2022, from [https://en.wikipedia.org/wiki/Transformer_\(machine_learning_model\)](https://en.wikipedia.org/wiki/Transformer_(machine_learning_model))

[24] Giacaglia, G. (2022, July 7). Transformers. Medium. Retrieved December 9, 2022, from <https://towardsdatascience.com/transformers-141e32e69591>

[25] Ankit, U. (2021, September 15). Transformer neural network: Step-by-step breakdown of the beast. Medium. Retrieved December 9, 2022, from <https://towardsdatascience.com/transformer-neural-network-step-by-step-breakdown-of-the-beast-b3e096dc857f>

[26] Dotti, P., & IT, D. tecnica A. (2022, January 27). Unsupervised learning O apprendimento non supervisionato. AI4Business. Retrieved December 10, 2022, from <https://www.ai4business.it/intelligenza-artificiale/unsupervised-learning-cose-ed-esempi-di-apprendimento-non-supervisionato/>

[27] Word2vec. Iampica.com. (n.d.). Retrieved December 12, 2022, from <https://readitaliano.com/wiki/it/Word2vec>

[28] MarketMuse. (2022, May 4). Topic modeling with word2vec. MarketMuse Blog. Retrieved December 12, 2022, from <https://blog.marketmuse.com/topic-modeling-with-word2vec-revised/>

[29] Jeffares, A. (2019, November 19). K-means: A complete introduction. Medium. Retrieved December 13, 2022, from <https://towardsdatascience.com/k-means-a-complete-introduction-1702af9cd8c>

[30] Wikimedia Foundation. (2022, October 20). GPT-3. Wikipedia. Retrieved December 13, 2022, from <https://it.wikipedia.org/wiki/GPT-3>

[31] Dale, R. (2020, December 15). GPT-3: What's IT good for?: Natural language engineering. Cambridge Core. Retrieved December 13, 2022, from <https://www.cambridge.org/core/journals/natural-language-engineering/article/gpt3-whats-it-good-for/0E05CFE68A7AC8BF794C8ECBE28AA990>

[32] OpenAI. (2022, December 12). CHATGPT: Optimizing language models for dialogue. OpenAI. Retrieved December 14, 2022, from <https://openai.com/blog/chatgpt/>

[33] Horev, R. (2018, November 17). Bert explained: State of the art language model for NLP. Medium. Retrieved December 14, 2022, from <https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270>

[34] Altobello, G. (2022, August 3). Cos'è bert, l'algoritmo che cambia il mondo del natural language processing. AI4Business. Retrieved December 14, 2022, from <https://www.ai4business.it/intelligenza-artificiale/cose-bert-lalgoritmo-che-cambia-il-mondo-del-natural-language-processing/>

[35] Budhiraja, A. (2018, March 6). Learning less to learn better-dropout in (deep) machine learning. Medium. Retrieved January 2, 2023, from <https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5>

[36] Blog: Quanti Dati Sono Generati Ogni minuto nel 2021? Info Data. (2022, July 1). Retrieved January 3, 2023, from https://www.infodata.ilsole24ore.com/2021/12/27/quant-dati-generati-minuto-nel-2021/?refresh_ce=1

[37] Google. (n.d.). Analisi dei Dati: Le attuali sfide e opportunità Nel Settore del marketing. Google. Retrieved January 3, 2023, from <https://www.thinkwithgoogle.com/intl/it-it/strategie/dati-e-misurazione/analisi-dei-dati-le-attuali-sfide-e-opportunita-nel-settore-del-marketing/>

[38] Maurizio.emilio. (2022, December 19). Intelligenza artificiale, Deep Learning e machine learning: Quali Sono Le Differenze. Innovation Post. Retrieved January 3, 2023, from <https://www.innovationpost.it/tecnologie/intelligenza-artificiale-deep-learning-e-machine-learning-quali-sono-le-differenze/>

[39] Academic.oup.com. (n.d.). Retrieved January 3, 2023, from <https://academic.oup.com/jamia/article/18/5/544/829676>

[40] Alammr, J. (n.d.). The illustrated transformer. The Illustrated Transformer – Jay Alammr – Visualizing machine learning one concept at a time. Retrieved January 3, 2023, from <https://jalammr.github.io/illustrated-transformer/>

[41] Picurelli, L. (2022, September 26). ▷ cos'è gpt 3 e perché sta rivoluzionando l'ia? Yeeply. Retrieved January 3, 2023, from <https://it.yeeply.com/blog/gpt-3-rivoluzione-ia/>

