

Presenta: Matteo Tramontano
Relatrice: Elena L. Piccolomini

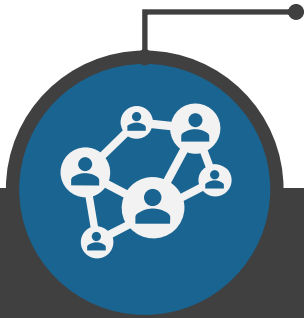
Sentiment Analysis

Previsione con tecniche di intelligenza artificiale

Sentiment Analysis

Web 2.0

Genericamente parlando, si intende un forte incremento dell'interazione tra sito e utente, con maggiore partecipazione dei fruitori che diventano anche autori



Big Data

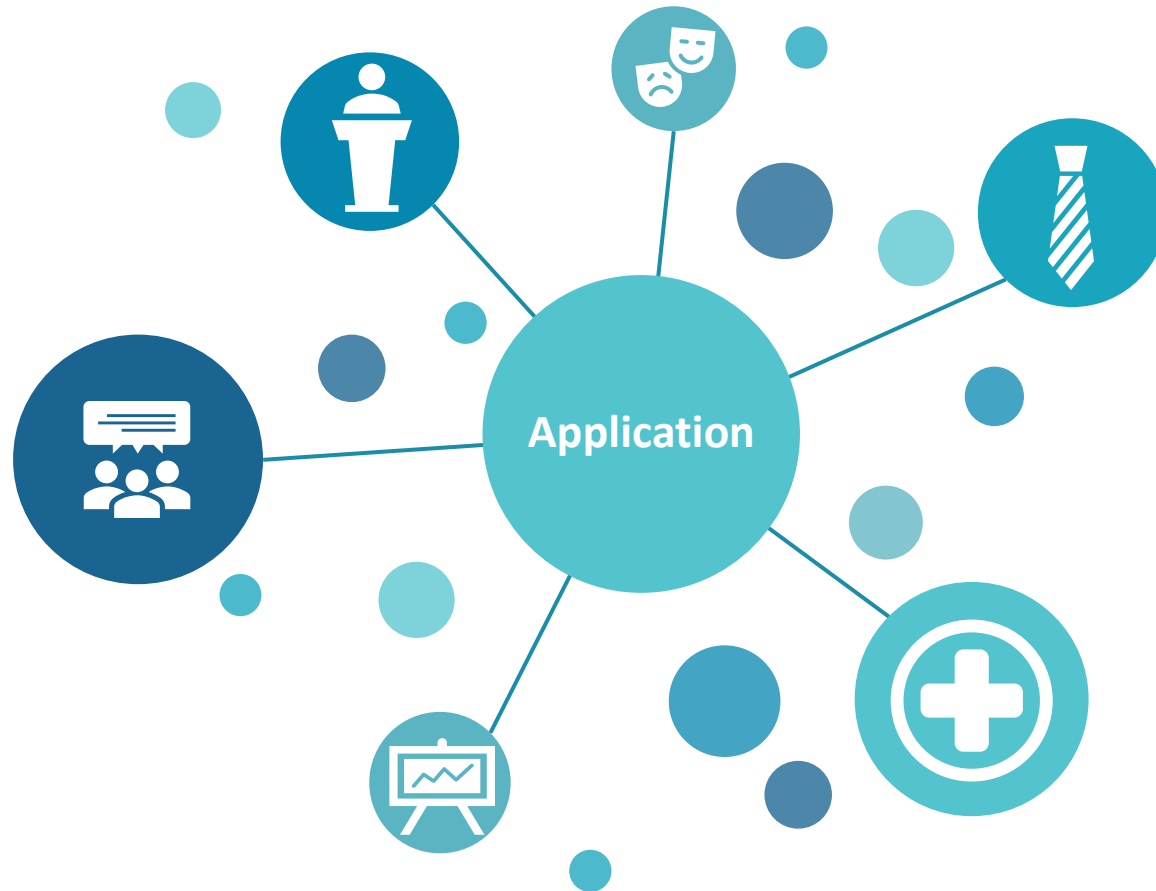
Raccolta di dati informatici così estesa in termini di volume, velocità e varietà da richiedere metodi analitici specifici per l'estrazione di valore o conoscenza.



Di cosa si tratta

Trattasi di un campo dell'elaborazione del linguaggio naturale (Natural Language Processing) che si occupa dell'identificazione ed estrazione di opinioni dal testo. Questa si basa sui principali metodi di linguistica computazionale e analisi testuale.

Applicazione e Scopo



Scopo Valutativo

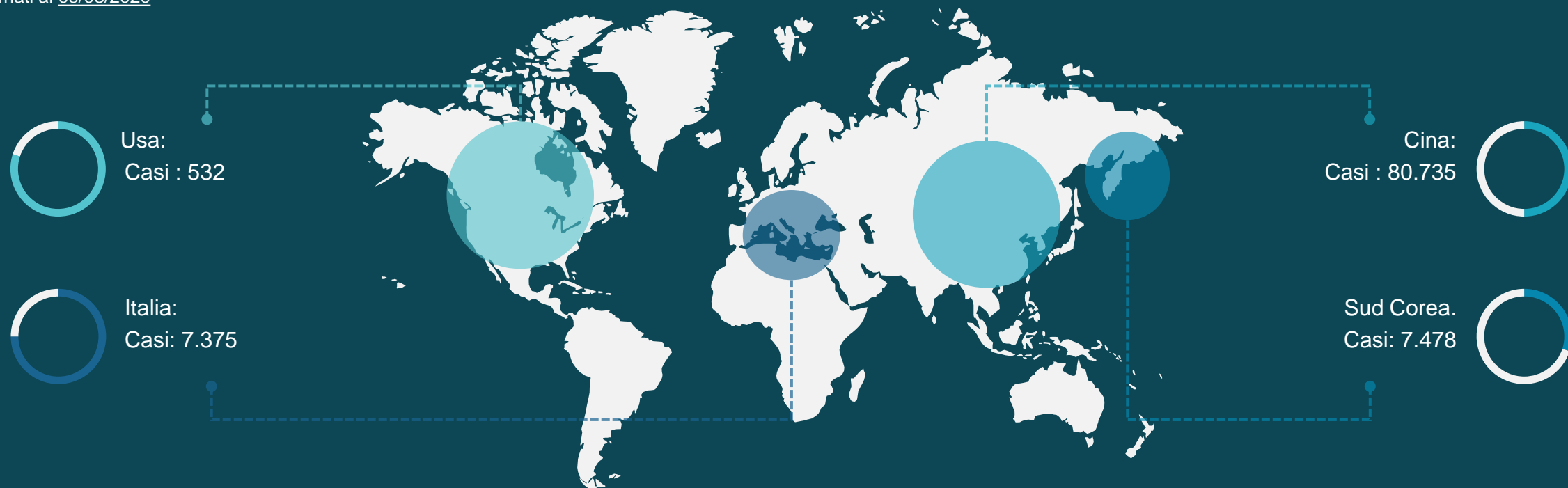
Utilizzata, principalmente dalle aziende, con lo scopo di effettuare valutazioni riguardo l'andamento di un qualsiasi suo aspetto. Reputazione, prodotti lanciati, assistenza, percezione del brand, misurare il ritorno delle attività di social media marketing...

Scopo Predittivo

Utilizzata al fine di effettuare previsioni riguardanti eventi futuri di qualsiasi tipo, la tecnica viene implementata in vari modi e può avere svariate applicazioni dal campo politico, di marketing, sportivo, medico...

“Blue Dot: la previsione del COVID-19”

Dati aggiornati al 09/03/2020



Bollettini Medici

Bollettini medici ufficiali



Compagnie Aeree

Dati delle compagnie aeree relative a prentoazioni e tratte



Malattie Epidemiche

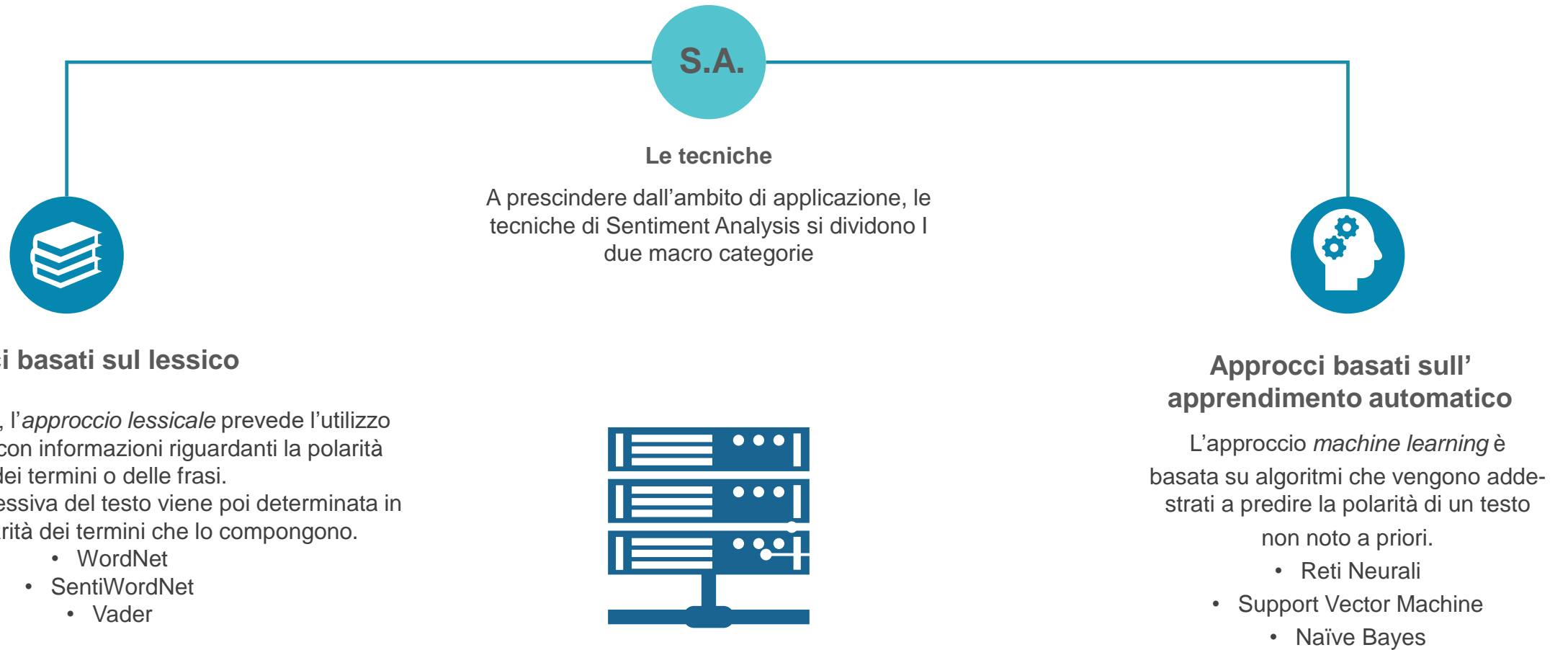
Segnalazione di malattie epidemiche anche degli animali.



65 Lingue Diverse

Il tutto in 65 differenti lingue.

Classificazione

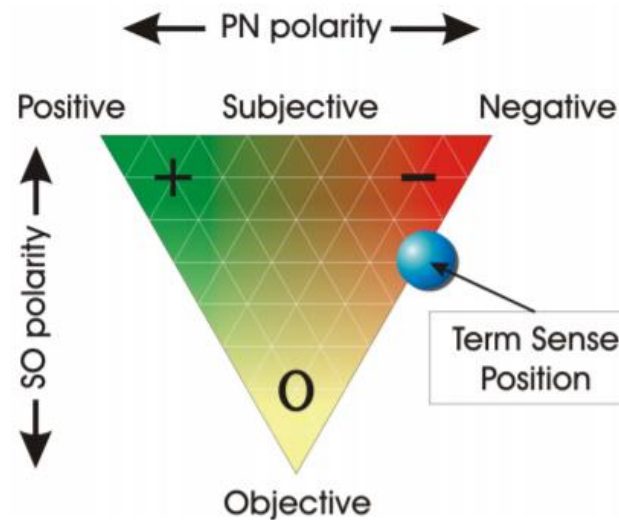


Lexicon Based Approach: *SentiWordNet*

SentiWordNet

Database semantico-lessicale frutto del genio di Andrea Esuli e Fabrizio Sebastiani con lo scopo di realizzare un database con cui effettuare S.A. basata sul lessico.

Questo dizionario applica ad ogni *synset* (gruppi di termini affini collegati tra loro tramite relazioni chiaramente definite) tre punteggi di polarità la cui somma è sempre uguale a 1. Tutti i termini dello stesso *synset* hanno quindi la stessa polarità e se un termine appartiene a più *synset* allora la sua polarità varia in base al contesto



PN-polarity

Indica il grado di positività/negatività del termine



SO-Polarity

Indica il grado di oggettività del termine



Pro

Nessun adattamento richiesto



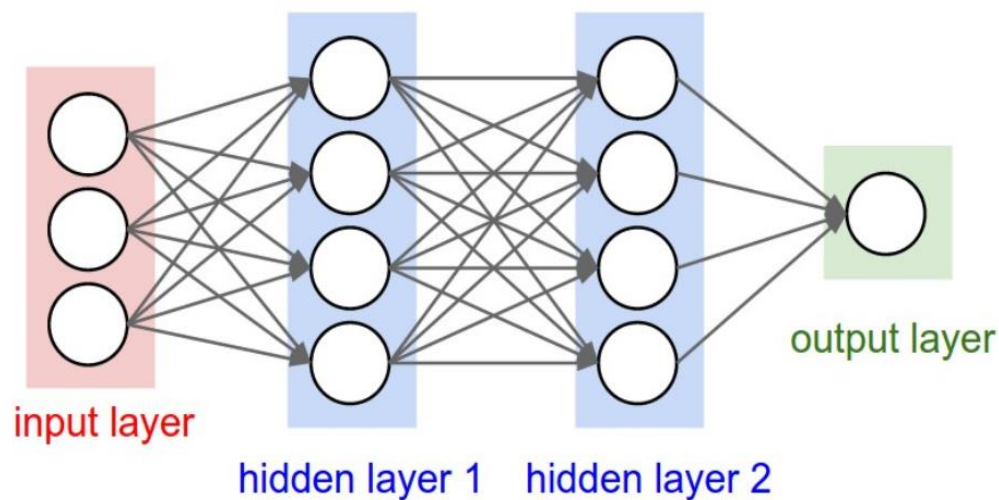
Contro

Basato sulla coerenza del lessico, non adattabile a specifici domini

Machine Learning approach: *deep learning*

Deep Learning

Rete neurale artificiale composta da almeno 2 livelli nascosti (solitamente molti di più), ad ogni connessione viene associato un peso, che determina l'importanza del valore di input (impostati inizialmente in maniera casuale), durante la valutazione alcuni dati hanno più importanza di altri, quindi tali connessioni avranno peso maggiore. Ogni neurone dispone di funzioni di attivazione con lo scopo di «standardizzare» l'output del neurone. Ogni volta che un insieme di dati attraversa tutti i layer della rete, restituisce il risultato tramite l'output layer.



Elevato parallelismo

Grazie al quale possono processare in tempi rapidi grandi quantità di dati



Tolleranza ai guasti

Grazie all'architettura parallela



Tolleranza al rumore

Riesce ad operare in modo corretto nonostante input imprecise o incompleti

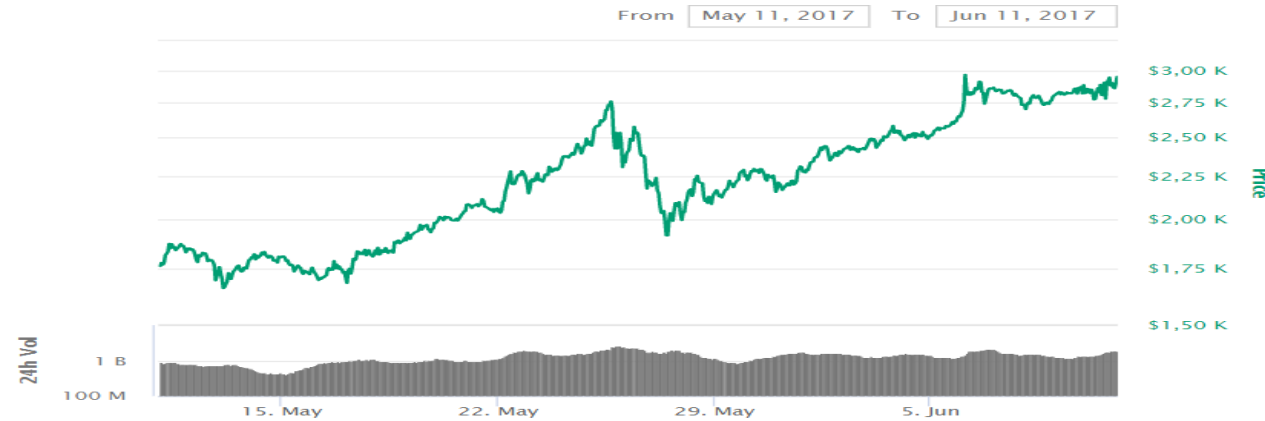


Evoluzione Adattiva

In grado (superata la fase di addestramento) di auto aggiornarsi in presenza di modifiche ambientali



Esempio di applicazione: «*Twitter & Bitcoin*»



Confusion Matrix	Predict Increase	Predict Decrease
Historical Increase	True Positive	False Negative
Historical Decrease	False Positive	True Negative

$$\text{Accuracy} = \frac{\sum(\text{TruePositive}) + \sum(\text{TrueNegative})}{\sum(\text{TotalPopulation})}$$

$$\text{Recall} = \frac{\sum(\text{TruePositive})}{\sum(\text{TruePositive}) + \sum(\text{FalseNegative})}$$

$$\text{Precision} = \frac{\sum(\text{TruePositive})}{\sum(\text{TruePositive}) + \sum(\text{FalsePositive})}$$

$$\text{F1-score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Risultati

Frequency	Accuracy	F1 score	Precision	Recall	Threshold
30min	0.787879	0.866667	0.722222	0.787879	2.25
45min	0.705882	0.700000	0.777778	0.736842	3.15
4h	0.661017	0.658537	0.818182	0.729730	0.20
2h	0.647059	0.777778	0.636364	0.700000	0.75
5min	0.630137	0.658537	0.675000	0.666667	9.10
15min	0.586207	0.777778	0.636364	0.700000	7.40



Ottimi risultati in presenza di determinate frequenza e punti di soglia, elevate Accuracy e Precision.



Particolari aspetti potrebbero avere influenzato l'analisi, come ad esempio l'andamento positivo del BTC nel periodo considerato.



THANK YOU

Matteo Tramontano

