

Puccini by mail

Sentiment polarity prediction

Andrea Ierardi

Contatti 

andrea.ierardi@student.unimi.it

andreierardi@gmail.com

Repository Github 

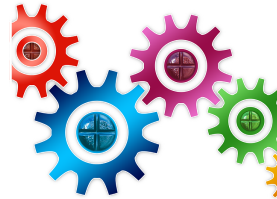
<https://github.com/Andreaierardi/sentiment-puccini>



Progetto



Sviluppo di modelli di previsione per rilevare la sentiment polarity



Addestramento dei modelli



Confronto tra modelli sviluppati e modelli pre-addestrati

Dati

Due dataset:

- Archivio Ricordi:
 - Estrazione lettere ricevute/mandate da/a Giacomo Puccini
- Sentipolc- evalita16:
 - Collezione di Tweet in lingua italiana
 - Disponibile sul sito in formato csv
 - 7000 esempi di training, 4000 esempi di test

Word Cloud

Le parole più utilizzate sono relative alla composizione, agli amici e ai collaboratori di Giacomo Puccini:

- **Composizioni:**

- Bohème, Manon Lescaut.

- **Amici e colleghi:**

- Tito Il Ricordi.
- Luigi Illica e Giuseppe Giacosa famosi librettisti con cui Puccini lavorò.
- Pietro Mascagni compagno di stanza durante il conservatorio.
- Leopoldo Mugnone.

- **Città italiane:**

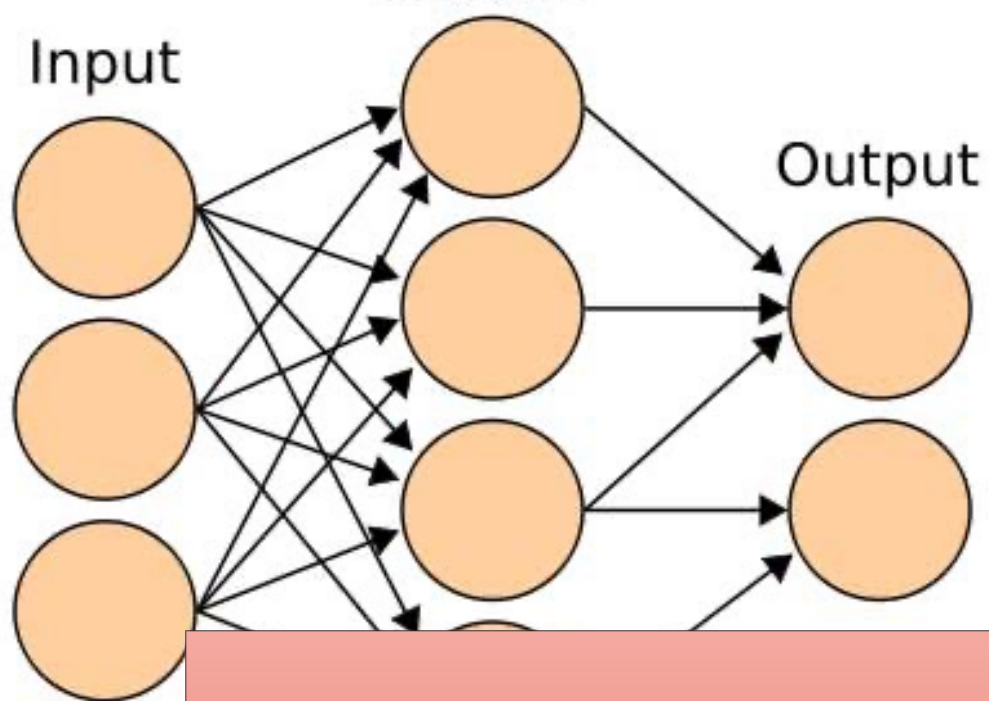
- Milano, Roma, Torino e Lucca
- Torre del Lago: una comunità nei pressi di Lucca dove Puccini si trasferì nel 1891.

- **Opera:**

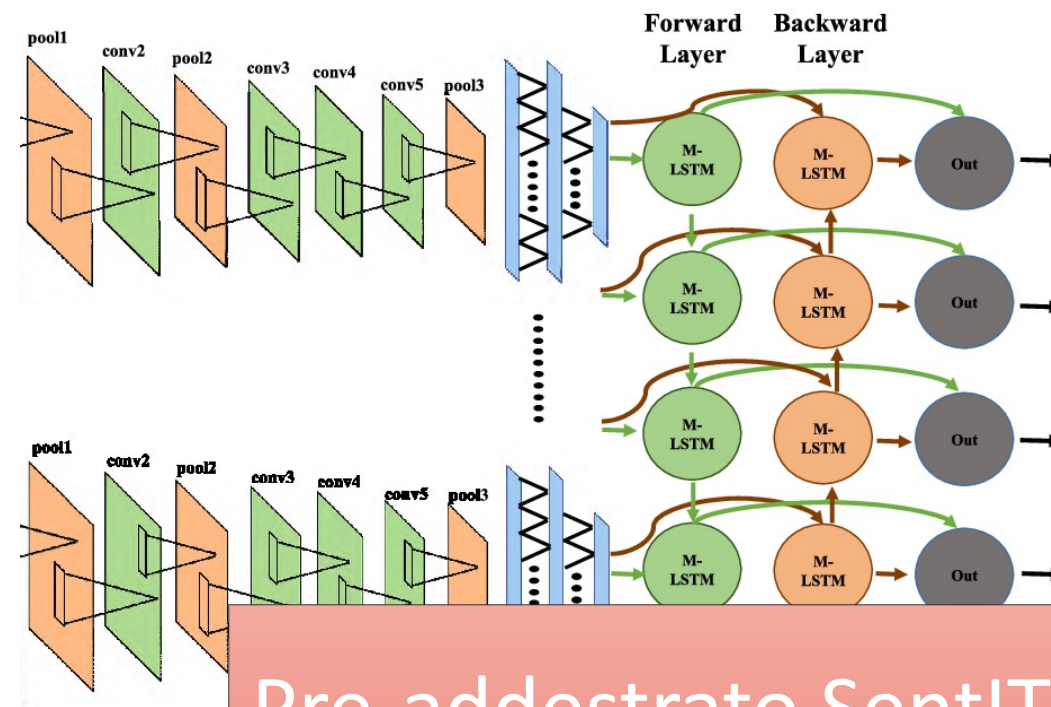
- Musica, versi, partitura, tempo, scena



Modelli



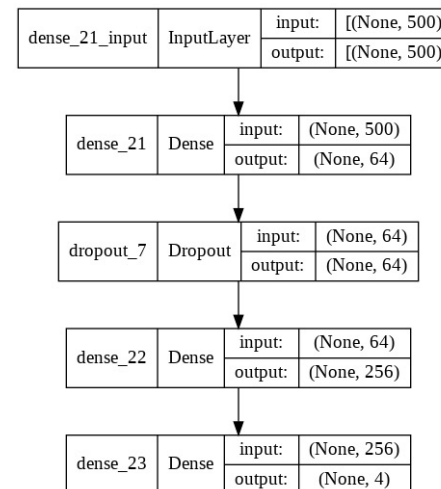
Reti Neurali Semplici



Pre-addestrato SentITA

Architettura delle reti neurali

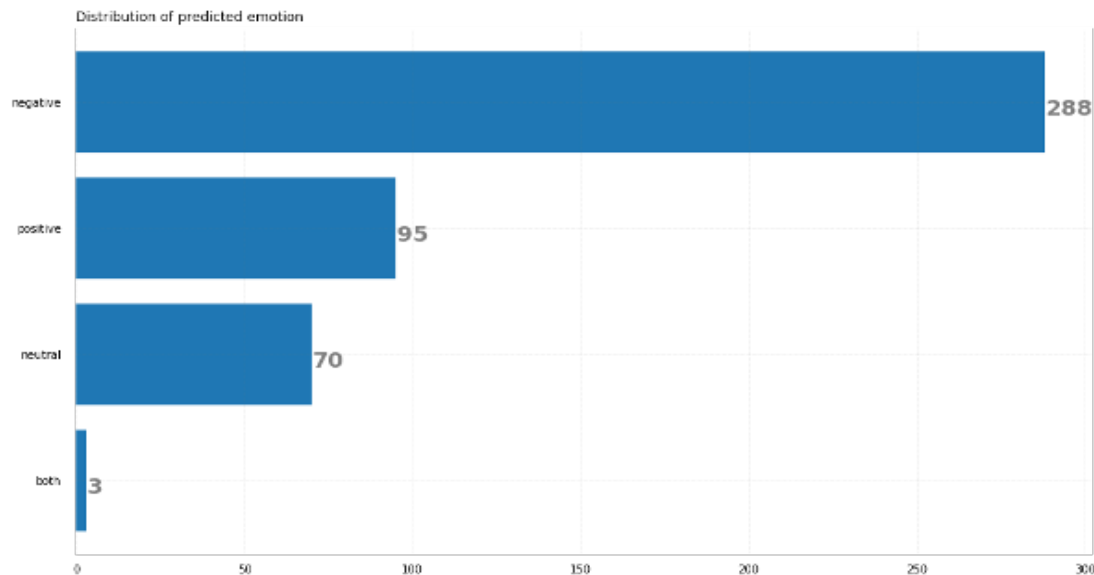
- Sono stati costruiti **due tipi** di Reti Neurali basate sul numero di differenti sentimenti della variabile target:
 - Rete Neurale a **2 sentimenti**: Positivi e Negativi
 - Rete Neurale a **4 sentimenti**: Positivi, Negativi, Entrambi e Neutrali.
- Architettura Rete Neurale Semplice



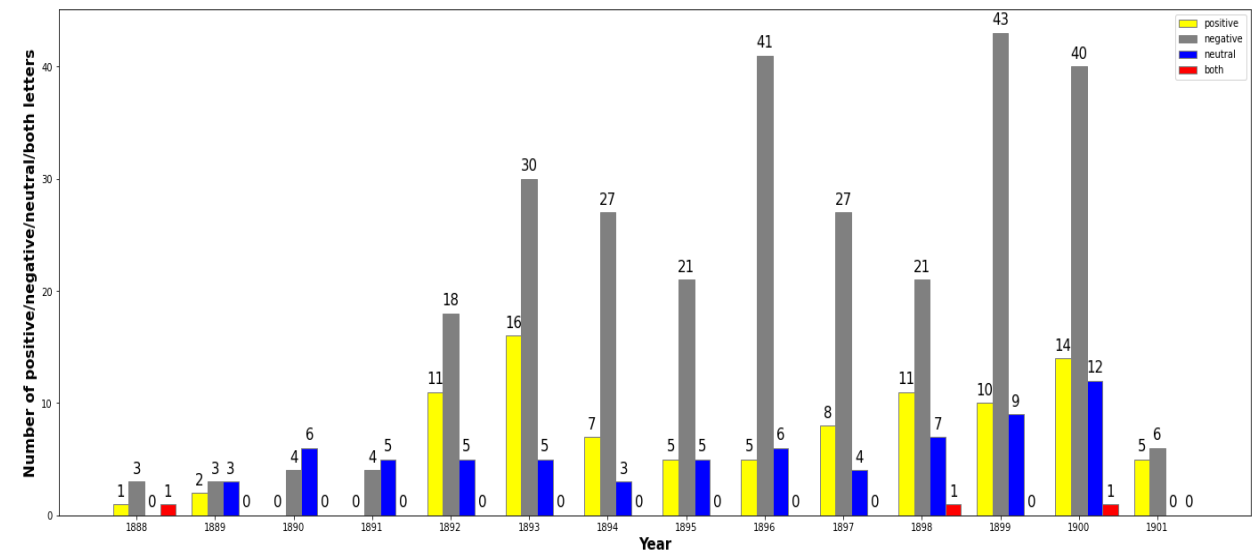
Reti Neurali a 4 sentimenti

- Risultati su SentiPolc Test set:
 - F1 score: 0.467 - Accuracy: 46.22 %
- Risultati modello sulle lettere di Puccini:

Previsioni modello a 4 sentimenti



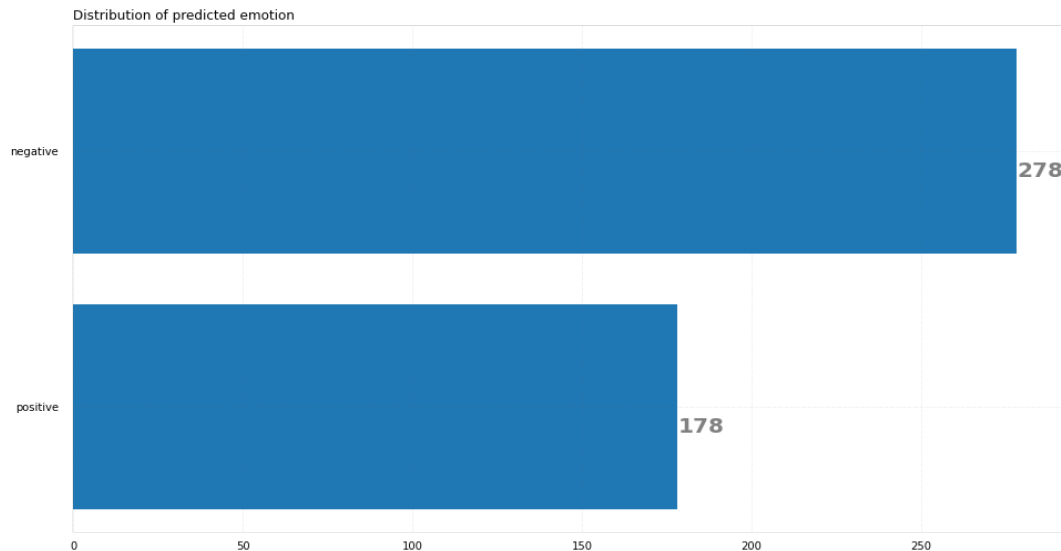
Previsioni modello a 4 sentimenti negli anni



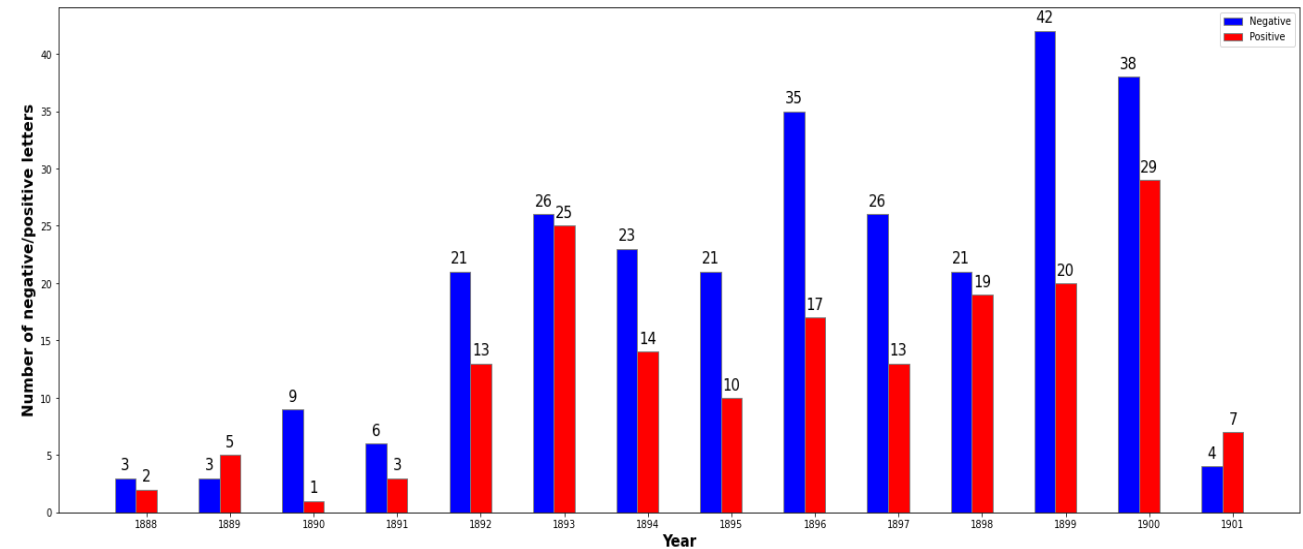
Reti Neurali a 2 sentimenti

- Risultati su SentiPolc Test set:
 - F1 score: 0.627 - Accuracy: 61.89 %
- Risultati modello sulle lettere di Puccini:

Previsioni modello a 2 sentimenti

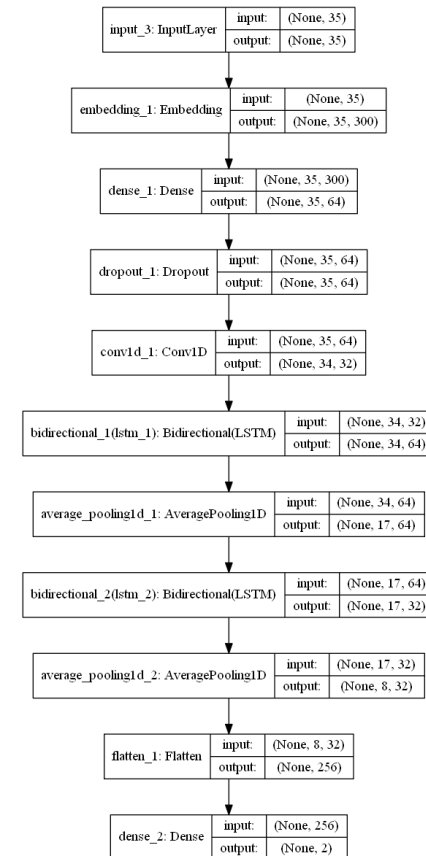


Previsioni modello a 2 sentimenti negli anni



Architettura SentITA

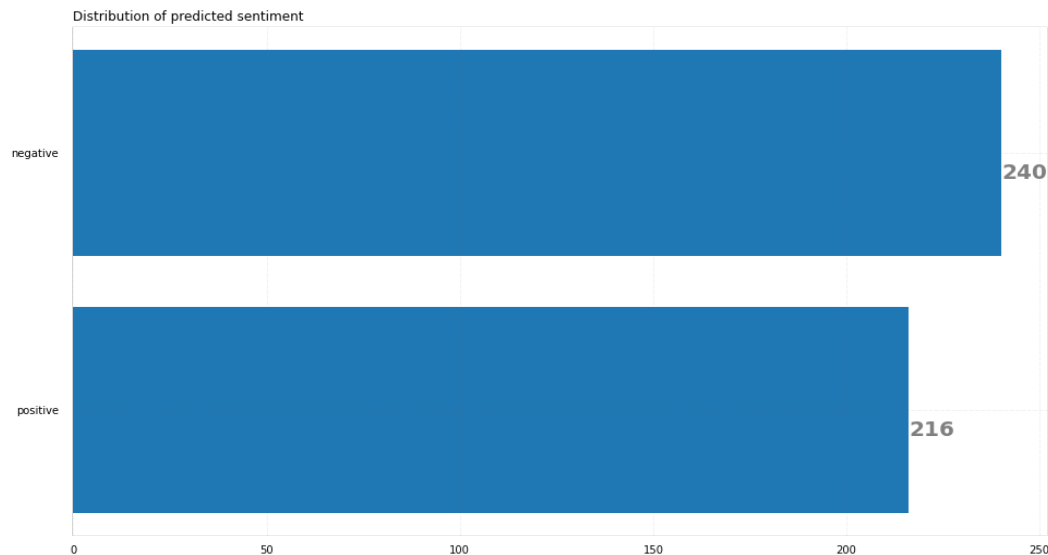
- Il modello riceve in input una rappresentazione delle parole con word embedding
- Addestrato su dataset differenti (Sentipolc2016, ABSITA2018 + Wikipedia).
- I dati di addestramento e test comprendono circa 102k frasi di cui 7k con sentimento positivo, 7k negativo e 88k neutrale.
- Architettura Rete Neurale: Bidirectional LSTM-CNN



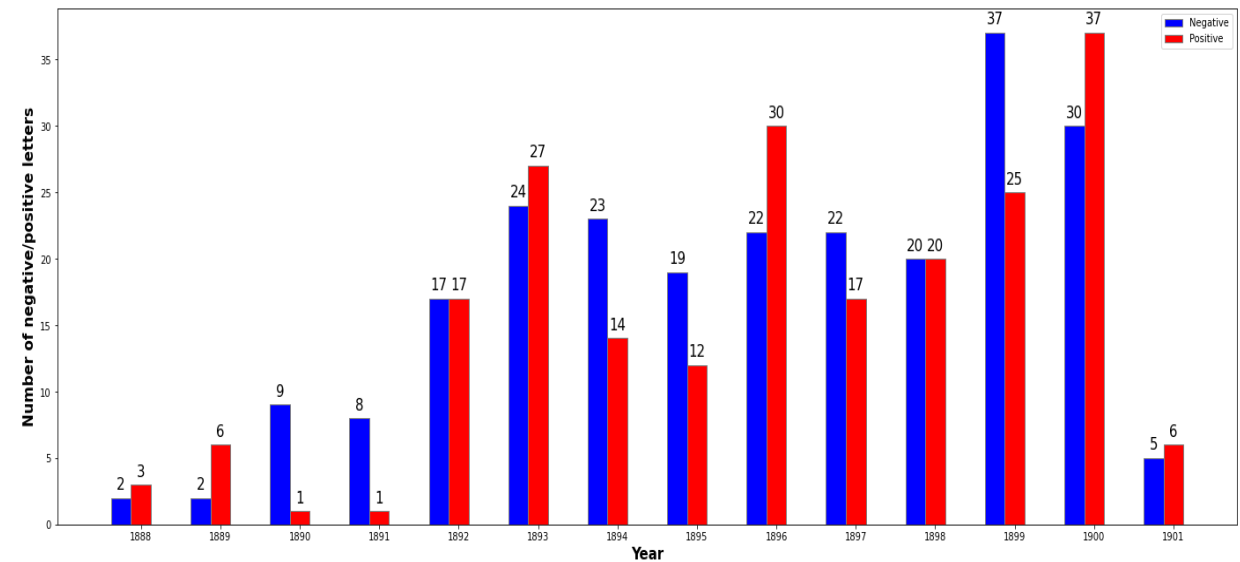
SentITA sentiment polarity

- Risultati su SentiPolc Test set:
 - F1 score: 0.85
- Risultati modello sulle lettere di Puccini:

Previsioni modello SentITA



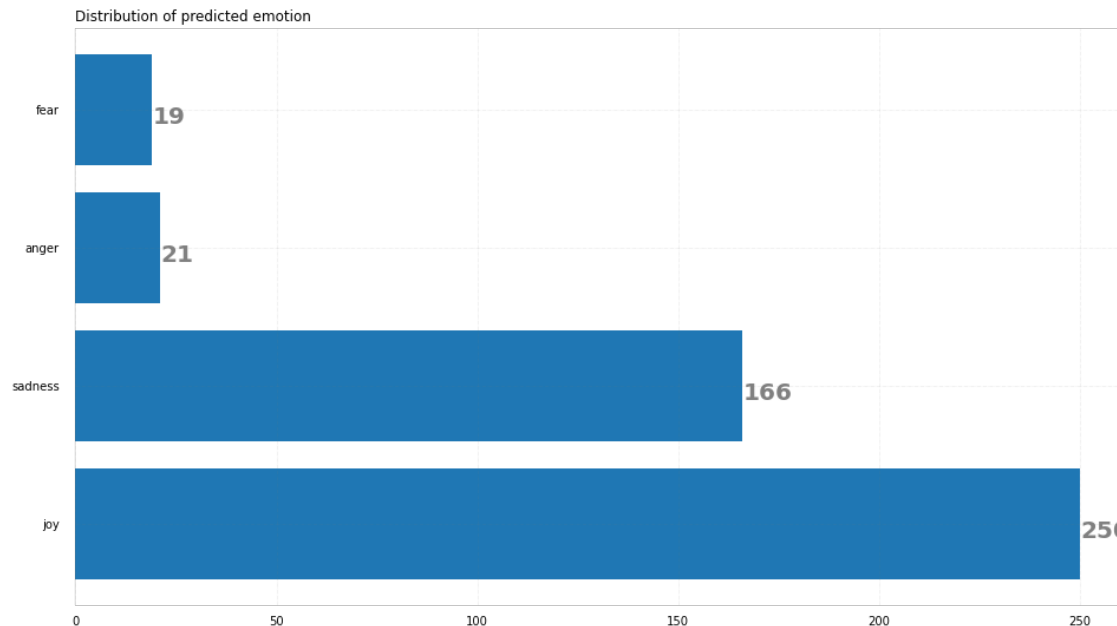
Previsioni modello SentITA negli anni



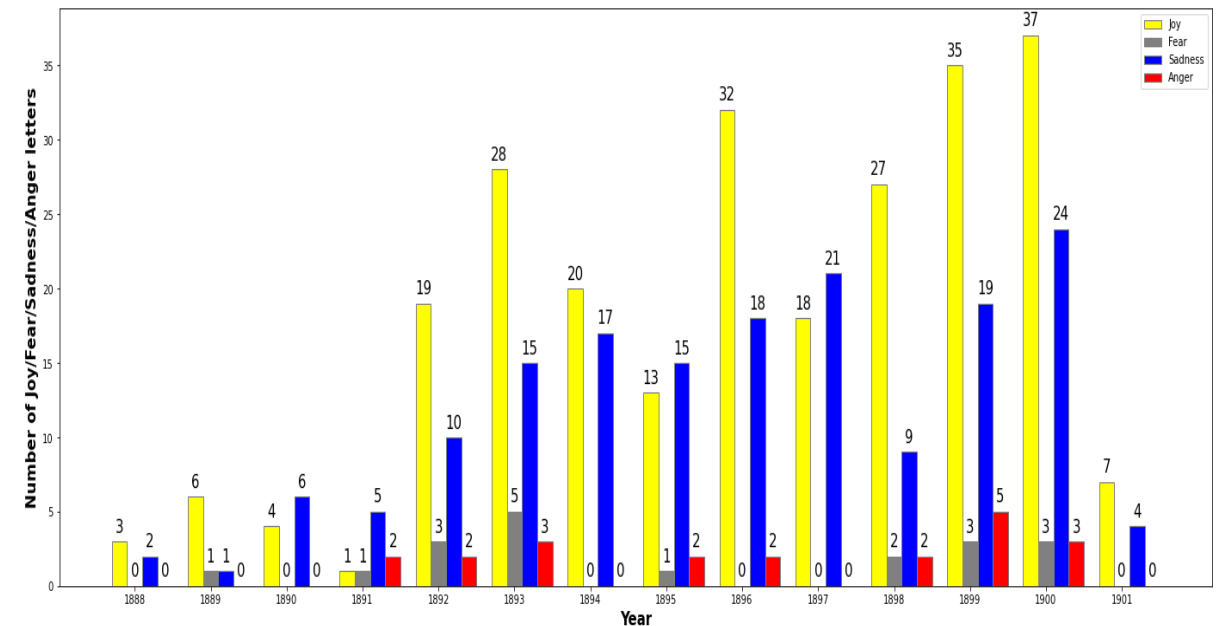
SentITA emotion

- Non ci sono informazioni sulle emozioni nei dati SentiPolc
- Risultati del modello sulle lettere di Puccini:

Previsioni emozioni modello SentITA



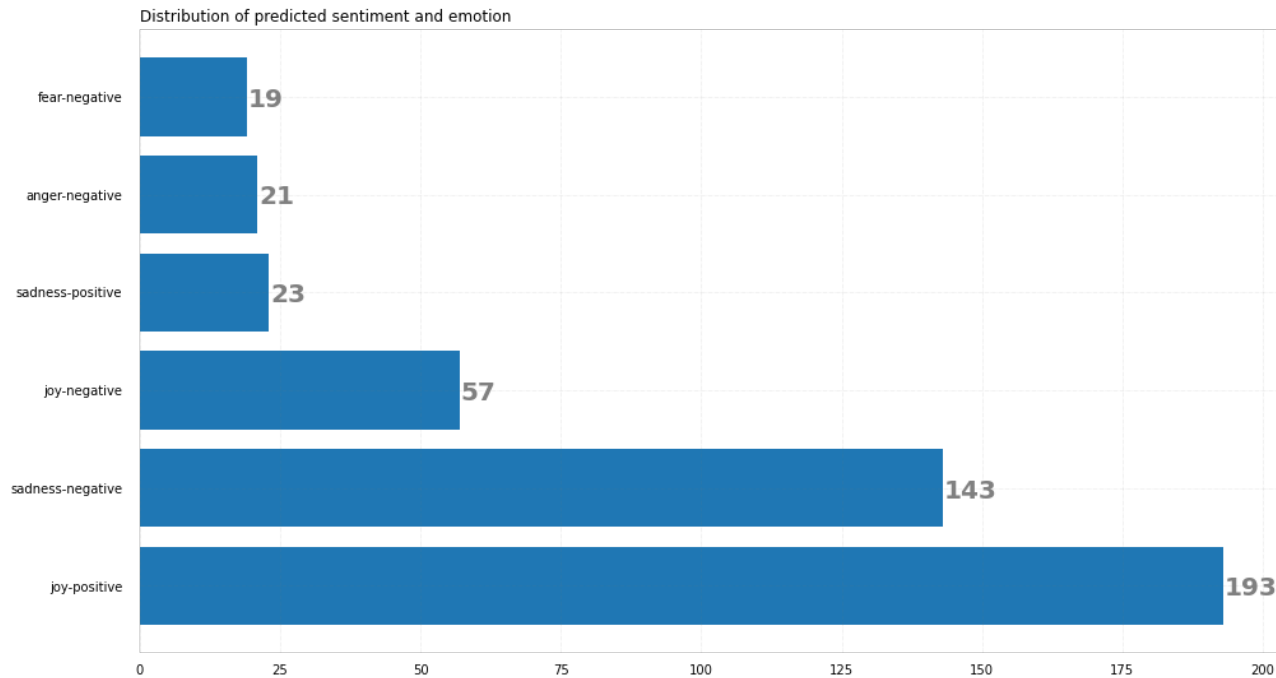
Previsioni emozioni modello SentITA negli anni



SentITA emotion and sentiment

- Non ci sono informazioni sulle emozioni nei dati SentiPolc
- Risultati modello sulle lettere di Puccini:

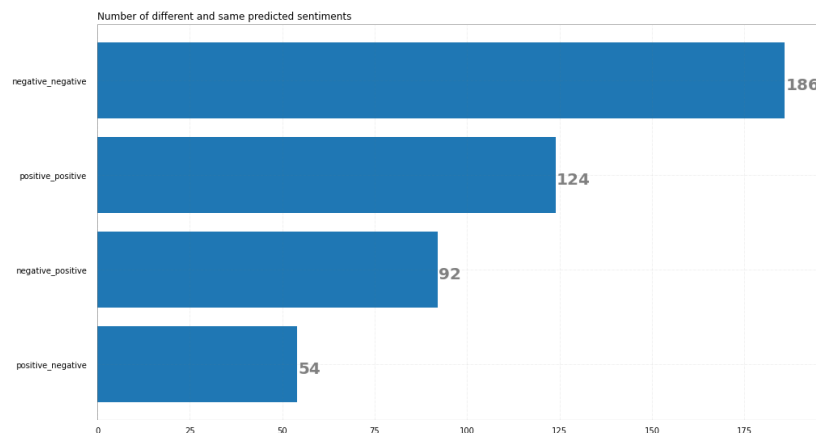
Previsioni emozioni-sentimenti modello SentITA



Neural Networks vs. SentITA

- Confronto tra una semplice Rete Neurale e il modello SentITA
- Per il confronto sono state considerate le sole previsioni relative alla sentiment polarity
- Confronto dei risultati sulle lettere di Puccini:

Confronto modello 2 sentimenti vs. SentITA

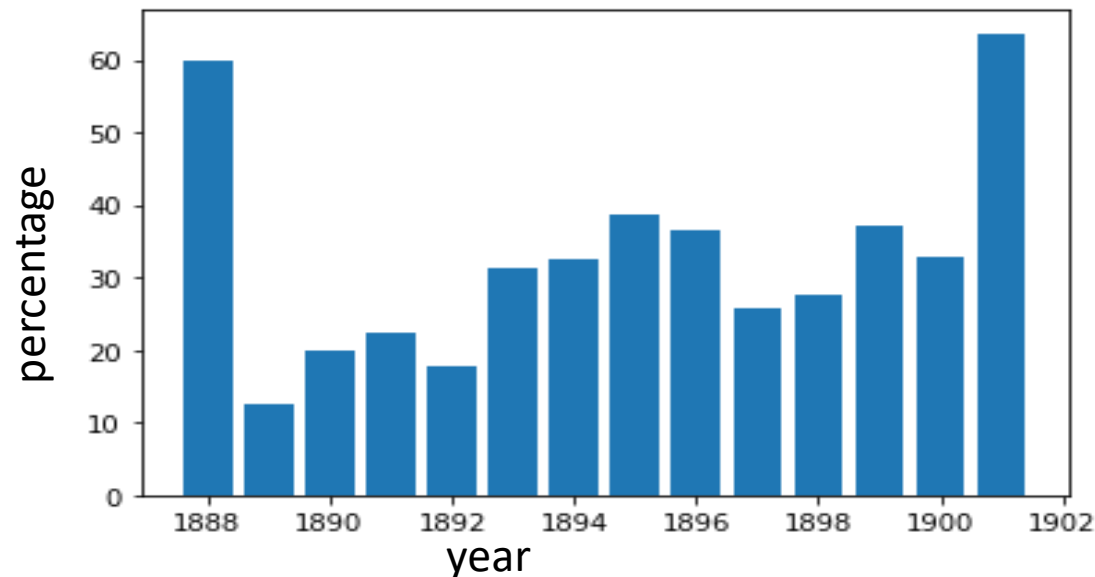


- La maggior parte delle lettere è stata classificata con lo stesso sentimento dai due modelli
- Il 68% delle lettere è classificato con lo stesso sentimento

Neural Networks vs. SentITA

- Confronto tra una semplice Rete Neurale e il modello SentITA
- Confronto risultati dei modelli sulle lettere di Puccini:

Percentuale di discordanza nelle previsioni tra Rete Neurale e SentITA



- La maggior parte delle lettere è distribuita negli anni centrali
- La percentuale è più alta nel 1888 e 1902 poiché il numero di lettere in quell'anno è basso
- La percentuale di discordanza rimane sotto il 40% negli altri anni.

Esempi di previsione

Milano 19 Giugno 1890,

Sigr. Marco Praga .

Vedendo che scorsero giorni e mesi, senza che ancora si abbia un risultato effettivo del lavoro di cui diedi incarico all'ottimo Puccini , e ciò **con danno rilevante** di tutte le parti interessate, maestro, poeti ed editore, ho dovuto persuadermi che il compositore si deve trovare innanzi a molte difficoltà che gli rendono impossibile il cammino. Ho voluto quindi farmi un concetto esatto intorno a tali **difficoltà**, e lasciando da parte quelle che concernono il dramma in generale, ed anche quelle riflettenti il 3° Atto, ho letto, riletto, studiato i primi due atti, dei quali il Puccini ha già musicato qualche parte. Ebbene, coscienziosamente ho dovuto convincermi della impossibilità di poterli musicare così come stanno: questi primi due atti verrebbero a durare musicalmente, ciò che invece l'opera intera richiederebbe di tempo, e perché risultino due atti regolari d'opera occorre un taglio cesareo, riducendoli alla metà circa. I raccorciamenti già praticati al 1° Atto, sono insufficienti affatto, e così com'è, senza ripetizioni di parole, sarebbero appena bastanti oltre 45 minuti di musica. Ora la durata giusta degli atti è uno dei primi elementi di successo: tanto più qui necessaria, perché si ha per le mani un dramma intimo, senza quell'ajuto di spettacolo grandioso che chiama l'attenzione del pubblico anche durante un lungo scorrere del tempo.

Affidandomi quindi alla mia lunga pratica, credo bene nell'interesse generale, e perché le cose camminino veloci, di occuparmi io stesso della riduzione dei due atti, progettando [sic] anche qualche cosa di nuovo pel 3°. Non parlo del 4° Atto perché non lo conosco ancora. Mostrerò poi a Lei ed all'eg° suo collega Avv. Oliva quanto ho creduto di fare, ed oso sperare avrà la loro piena approvazione. Così avremo tutti raggiunto lo scopo, che è quello di vedere l'opera presto compiuta e pronta per le scene.

Ho creduto comunicarle ciò confidenzialmente, onde spiegarle anche il mio silenzio prolungato. Spero avere il piacere di vederla presto, ed intanto con tutta stima mi ripeto

devtmo Suo Giulio Ricordi

2-sentiment
prediction
negative

SentITA sentiment
prediction
negative

4-sentiment
prediction
negative

SentITA emotion
detection
sadness

Esempi di previsione

Puccini.

Teatro Regio. Torino

Fuori casa ebbi tardi telegramma. Sono sbalordito tonto più non sapendo come rimediare non essendovi artisti appena possibili. Qualora confermasi **deficiente** non rimarrebbe che ritirare opera. Sono ora impegnato ma occorrendo farà corsa entro settimana. Intanto provi ancora con pazienza sembrandomi quasi impossibile artista ebbe sempre successo manchi ora completamente. Saluti.

Ricordi

2-sentiment
prediction
negative

SentITA sentiment
prediction
negative

4-sentiment
prediction
negative

SentITA emotion
detection
anger

Esempi di previsione

17 del 93

Carissimo Puccini,

Ricevo la **gradita** tua cartolina d'jeri e d'incarico del Sr Com. Giulio ti mando qui accluso un Bono Banca Nazle di £200-.

Il M° Mattaress ti ringrazia della tua premura e ti saluta caramente.

Non puoi immaginarti come noi siamo tutti **contenti**, sapendoti soddisfatto delle prove d'orchestra Manon!

Coi saluti cordiali dell' eg° ns Sr Tornaghi, m'è caro ripetermi

tuo aff° amico

C. Blanc

Salutami Carignani

2-sentiment
prediction
positive

SentITA sentiment
prediction
negative

4-sentiment
prediction
negative

SentITA emotion
detection
joy

Esempi di previsione

Puccini

Teatro Regio. Torino

Lietissimo **buone** notizie-**Evviva**. Spedisco messa scena. Cominciate sbizzare prove su questa. Verrò Domenica sera 10 e 47. Disponete due buone prove scena Lunedì non potendo fermarmi oltre. A rivederci.

Ricordi

2-sentiment
prediction
negative

SentITA sentiment
prediction
positive

4-sentiment
prediction
positive

SentITA emotion
detection
joy

Conclusioni & Next Steps

- Conclusioni

- Il modello SentITA ha dato risultati migliori ma la grandezza dei dataset di addestramento rende il procedimento di previsione lento. Infatti, la Rete Neurale a 2 sentimenti è parecchio più veloce da addestrare e potrebbe quindi essere impiegata come modello baseline affidabile.

- Next Steps:

- Sviluppo di modelli semplici anche per la parte di emotions detection
- Utilizzo di altre variabili per l'addestramento dei modelli (emittente, destinatario, data, luogo, volume, anno, pagine ecc.)

Domande?

