



UNIVERSITÀ DEGLI STUDI DI MILANO - BICOCCA

**Scuola di Scienze**

**Dipartimento di Informatica, Sistemistica e Comunicazione**

**Corso di laurea in Informatica**

# **Enhancing irony detection with affective information**

**Relatore:** Prof. Elisabetta Fersini

**Relazione della prova finale di:**

Gianluca Giudice

Matricola 830694

Anno Accademico 2019-2020

# Contents

<b>1</b>	<b>Stato dell'arte</b>	<b>3</b>
1.1	Approccio supervisionato . . . . .	3
1.2	Approccio non supervisionato . . . . .	3
1.3	Approccio semi supervisionato . . . . .	3
<b>2</b>	<b>Sistema realizzato</b>	<b>4</b>
2.1	Descrizine del sistema proposto . . . . .	4
2.2	Rappresentazione del testo . . . . .	4
2.2.1	Rappresentazione booleana . . . . .	4
2.2.2	Rappresentazione mediante transformer . . . . .	4
2.3	Caratteristiche linguistiche . . . . .	4
2.3.1	PP . . . . .	4
2.3.2	POS . . . . .	4
2.3.3	Onomatopeic . . . . .	4
2.3.4	...Tutto il resto . . . . .	4
2.3.5	EMOT . . . . .	4
2.4	Modelli supervisionati . . . . .	4
2.4.1	Alberi di decisione . . . . .	4
2.4.2	Support Vector Machine . . . . .	4
2.4.3	Naive Bayes . . . . .	4
2.4.4	Bayesian Network . . . . .	4
2.5	Strumenti utilizzati . . . . .	4
2.5.1	Scikit-learn . . . . .	4
2.5.2	Weka . . . . .	4
<b>3</b>	<b>Campagna sperimentale</b>	<b>5</b>
3.1	Dataset . . . . .	5
3.2	Misure di performance . . . . .	5
3.3	BOW + Caratteristiche linguistiche . . . . .	5
3.4	BERT + Caratteristiche linguistiche . . . . .	5
3.5	SBERT + Caratteristiche linguistiche . . . . .	5
3.6	Analisi lessico con PCA . . . . .	5
<b>4</b>	<b>Conclusioni e sviluppi futuri</b>	<b>6</b>

# Introduzione

Descrizione del problema

Approccio al problema

Sintesi dei risultati

# Chapter 1

## Stato dell'arte

1.1 Approccio supervisionato

1.2 Approccio non supervisionato

1.3 Approccio semi supervisionato

## Chapter 2

# Sistema realizzato

### 2.1 Descrizine del sistema proposto

### 2.2 Rappresentazione del testo

#### 2.2.1 Rappresentazione booleana

#### 2.2.2 Rappresentazione mediante transformer

### 2.3 Caratteristiche linguistiche

#### 2.3.1 PP

#### 2.3.2 POS

#### 2.3.3 Onomatopeic

#### 2.3.4 ...Tutto il resto

#### 2.3.5 EMOT

### 2.4 Modelli supervisionati

#### 2.4.1 Alberi di decisione

#### 2.4.2 Support Vector Machine

#### 2.4.3 Naive Bayes

#### 2.4.4 Bayesian Network

### 2.5 Strumenti utilizzati

#### 2.5.1 Scikit-learn

#### 2.5.2 Weka

## Chapter 3

# Campagna sperimentale

3.1 Dataset

3.2 Misure di performance

3.3 BOW + Caratteristiche linguistiche

3.4 BERT + Caratteristiche linguistiche

3.5 SBERT + Caratteristiche linguistiche

3.6 Analisi lessico con PCA

## Chapter 4

# Conclusioni e sviluppi futuri