

# L'Epistola a Cangrande al vaglio della *Authorship Verification*

Silvia Corbara♣, Alejandro Moreo◇, Fabrizio Sebastiani◇,  
Mirko Tavoni♣

♣ Dipartimento di Filologia, Letteratura e Linguistica  
Università di Pisa  
56126 Pisa, IT

◇ Istituto di Scienza e Tecnologie dell'Informazione  
Consiglio Nazionale delle Ricerche  
56124 Pisa, IT

Pisa, 18 dicembre 2018  
Seminario di Studi  
“Nuove inchieste sull'Epistola a Cangrande”

# Introduzione alla (*Computational*) *Authorship Analysis*

- (**Computational**) **Authorship Analysis** (AA): metodi automatici per l'identificazione della *authorship* di un documento testuale di paternità ignota o controversa
- Vari *subtask* in cui la AA si declina:
  - **Authorship Verification** (AV): dato un candidato predefinito, dire se l'autore del documento è questo candidato o meno
  - **Authorship Attribution**: dato un insieme finito e predefinito di candidati, identificare quale di essi è l'autore del documento
  - **Shared Authorship**: dati due documenti, dire se sono o meno dello stesso autore
  - ...

# Introduzione alla (*Computational*) *Authorship Analysis*

- Aree di applicazione della AA:
  - Filologia
  - **Cybersecurity**, i.e., prevenzione di reati che potrebbero essere commessi con l'ausilio di mezzi digitali
  - **Computational forensics**, i.e., analisi di evidenze riguardanti reati già avvenuti
- In ciascun *subtask*, oltre a formulare un'ipotesi, queste tecnologie indicano anche il “grado di confidenza” che il sistema ha nell'ipotesi formulata
- L'indagine sull'EaC può essere definita come un caso di **adversarial authorship verification**: dato un candidato predefinito, dire se l'autore del documento è questo candidato o una persona **che si finge** il candidato

# Scopi di questo lavoro

- Nell'applicare tecniche di AV all'EaC abbiamo tre scopi:
  - ① Formulare un'ipotesi se Dante abbia scritto o meno (la prima e/o la seconda parte del)l'EaC

# Scopi di questo lavoro

- Nell'applicare tecniche di AV all'EaC abbiamo tre scopi:
  - ① Formulare un'ipotesi se Dante abbia scritto o meno (la prima e/o la seconda parte del)l'EaC
  - ② Minimizzare la percentuale di errore (**error rate**) dell'algoritmo con cui abbiamo compiuto la determinazione, calcolata su un insieme di testi di paternità certa
    - Per ciascuno di questi testi, fingiamo di non conoscerne la paternità e controlliamo se il sistema la indovina correttamente
    - L'ipotesi (1) è **credibile** tanto più bassa è questa error rate

# Scopi di questo lavoro

- Nell'applicare tecniche di AV all'EaC abbiamo tre scopi:
  - ① Formulare un'ipotesi se Dante abbia scritto o meno (la prima e/o la seconda parte del)l'EaC
  - ② Minimizzare la percentuale di errore (**error rate**) dell'algoritmo con cui abbiamo compiuto la determinazione, calcolata su un insieme di testi di paternità certa
    - Per ciascuno di questi testi, fingiamo di non conoscerne la paternità e controlliamo se il sistema la indovina correttamente
    - L'ipotesi (1) è **credibile** tanto più bassa è questa error rate
  - ③ Esibire i motivi che hanno portato a questa ipotesi, in modo tale che essi possano diventare nuove ipotesi di studio per il filologo
    - L'ipotesi (1) è **utile** se e solo se è accompagnata dalle evidenze che la supportano

# Scopi di questo lavoro

- Nell'applicare tecniche di AV all'EaC abbiamo tre scopi:
  - ① Formulare un'ipotesi se Dante abbia scritto o meno (la prima e/o la seconda parte del)l'EaC
  - ② Minimizzare la percentuale di errore (**error rate**) dell'algoritmo con cui abbiamo compiuto la determinazione, calcolata su un insieme di testi di paternità certa
    - Per ciascuno di questi testi, fingiamo di non conoscerne la paternità e controlliamo se il sistema la indovina correttamente
    - L'ipotesi (1) è **credibile** tanto più bassa è questa error rate
  - ③ Esibire i motivi che hanno portato a questa ipotesi, in modo tale che essi possano diventare nuove ipotesi di studio per il filologo
    - L'ipotesi (1) è **utile** se e solo se è accompagnata dalle evidenze che la supportano

## The Takeaway Message

L'algoritmo di AV non si vuole sostituire al filologo, ma gli vuole fornire nuove ipotesi di lavoro da validare

# Introduzione alla (*Computational*) *Authorship Analysis*

- L'AA si basa non sugli eventi descritti / concetti espressi nel documento (e/o la loro veridicità o plausibilità), ma sullo studio stilistico quantitativo (**stilometria**) del linguaggio con cui tali eventi vengono descritti / concetti vengono espressi
- Tentativo di catturare i tratti inconsci (“tic compositivi”) dell’uso del linguaggio (**writer invariants**)



# Introduzione alla (*Computational*) *Authorship Analysis*

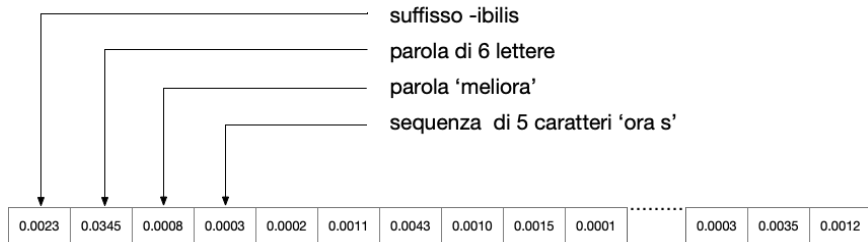
- L'AA si basa non sugli eventi descritti / concetti espressi nel documento (e/o la loro veridicità o plausibilità), ma sullo studio stilistico quantitativo (**stilometria**) del linguaggio con cui tali eventi vengono descritti / concetti vengono espressi
- Tentativo di catturare i tratti inconsci (“tic compositivi”) dell’uso del linguaggio (**writer invariants**)
- Tre fasi nella storia della disciplina
  - (– 1964): Identificazione basata su una singola caratteristica numerica del testo
  - (1964 – 1999): Identificazione basata su caratteristiche numeriche multiple del testo
  - (2000 –): Identificazione basata su metodi di *text classification* e *machine learning*
- Le prime due fasi si distinguono fra di loro per il numero (e il tipo) delle caratteristiche da analizzare, la terza si distingue dalle prime due per il tipo di modelli matematici usati per analizzare le caratteristiche prescelte

# Cos'è il *Machine Learning*?

- **Machine Learning** : la sottodisciplina dell'informatica che si occupa di algoritmi che apprendono a risolvere un determinato problema (e.g. *authorship verification*) da un insieme di esempi correttamente risolti di questo problema (*training examples*)
- **Classificazione** : assegnazione di oggetti a classi predefinite; è la “madre di tutti i problemi” risolvibili mediante ML
- La *authorship verification* per l'EaC è un problema di **classificazione binaria**, dato che consiste nell'assegnare un documento di paternità ignota a una delle due classi {Dante, NonDante}
- Ciò richiederà di “allenare” il classificatore esponendolo
  - a un insieme di testi attribuiti con certezza a Dante (**positive training examples**)
  - a un insieme di testi attribuiti con certezza ad autori diversi da Dante (ma a lui “simili”) (**negative training examples**)

# Machine Learning per la Authorship Verification

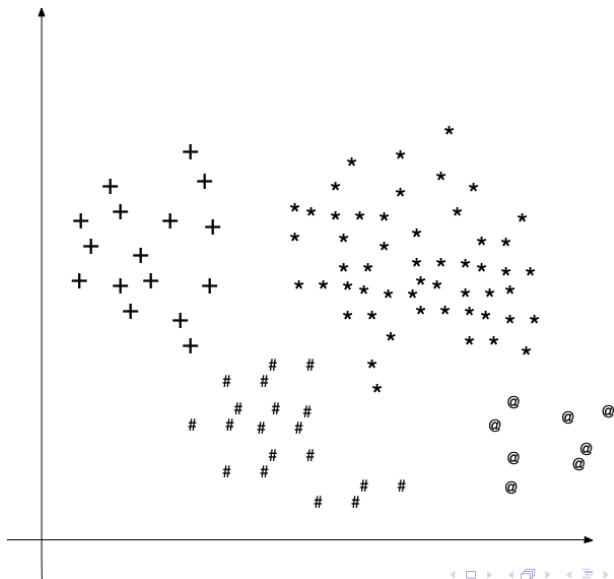
- Un algoritmo di machine learning richiede che ciascun testo (i.e., un testo di training, o il testo da attribuire) vada rappresentato come un insieme (**vettore**) di caratteristiche (**feature**)
- Nella AA, ogni *feature* rappresenta la **frequenza di occorrenza** nel documento di un certo tratto stilistico



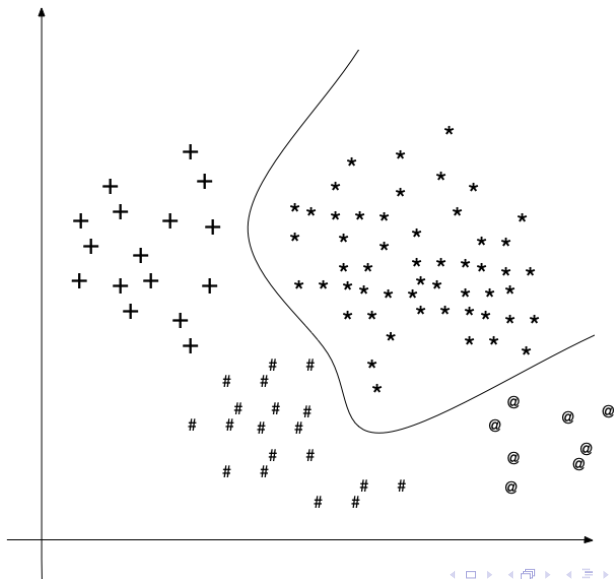
# Machine Learning per la Authorship Verification

- *Feature* tipiche della AA sono
  - parole
  - *character n-grams*
  - suffissi
  - segni di interpunzione
  - lunghezze delle parole
  - lunghezze delle frasi
  - *POS-tags*
  - ...
- L'ipotesi è che
  - due testi dello stesso autore siano rappresentati da vettori “simili”
  - due testi di diverso autore siano rappresentati da vettori “dissimili”
- I vettori possono contenere anche decine di migliaia di *feature*
- È l'algoritmo stesso a capire (dai testi di training) quali caratteristiche sono “discriminanti” e quali no

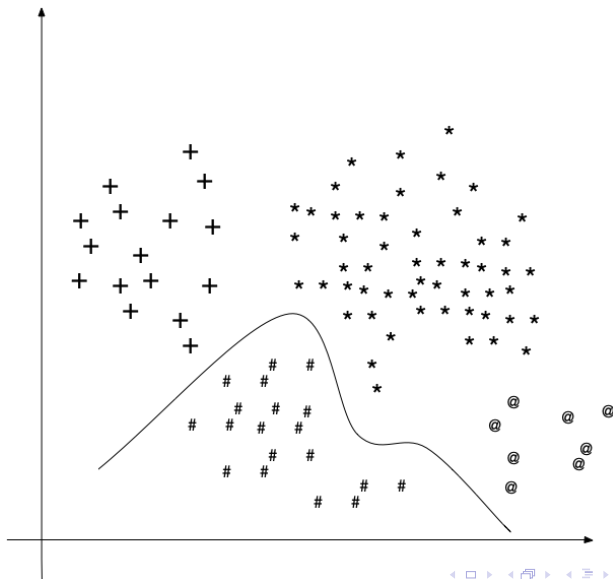
# Un'interpretazione geometrica della AV



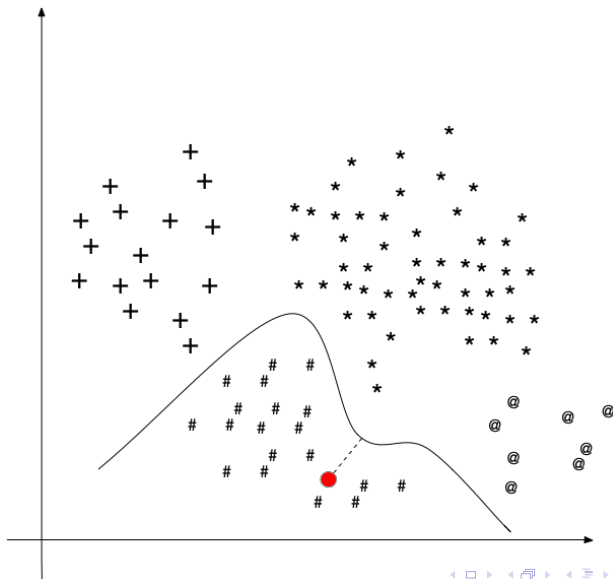
# Un'interpretazione geometrica della AV



# Un'interpretazione geometrica della AV



# Un'interpretazione geometrica della AV





## Training examples usati in questo lavoro

### PARTE NUNCUPATORIA DELLA EAC (EPISTOLA I)

Autore	Opere
Dante	<i>Epistole</i>
Vari autori Giovanni Boccaccio	Dalla raccolta di epistole cancelleresche di Petrus de Boateriis <i>Epistole e lettere</i>

### PARTE ESEGETICA DELLA EAC (EPISTOLA II)

Autore	Opere
Dante	<i>De Vulgari Eloquentia</i> <i>Monarchia</i>
Giovanni Del Virgilio Graziolo Bambaglioli Guido Da Pisa Zono de' Magnalis Giovanni Boccaccio	<i>Allegorie super fabulas Ovidii Methamorphoseos</i> <i>Comento all'Inferno di Dante</i> <i>Expositiones et glose. Declaratio super Comediam Dantis</i> <i>Vita di Virgilio</i> <i>De vita et moribus d. Francisci Petracchi</i> <i>De Mulieribus Claris</i>
Pietro Alighieri Benvenuto Da Imola	<i>Comentum super poema Comedie Dantis</i> <i>Comentum super Dantis Aldigherij Comoediam</i> <i>Glose Bucolicorum Virgilii</i> <i>Expositio super Valerio Maximo</i>
Filippo Villani	<i>Expositio seu comentum super Comedia Dantis Allegherii</i>

## Training examples usati in questo lavoro

#Documenti	EPISTOLA I	EPISTOLA II
Dante	12	2
NonDante	55	11

#Parole	EPISTOLA I	EPISTOLA II
Dante	6.066	30.578
NonDante	31.130	538.994

#Caratteri	EPISTOLA I	EPISTOLA II
Dante	43.832	210.309
NonDante	219.033	3.616.915

# Il Classificatore

- L'algoritmo di *machine learning* è una *Support Vector Machine* (SVM)
- Le *feature* che abbiamo al momento:
  - *character n-grams* di lunghezza 3, 4 e 5 (6.678)
  - parole funzione (77)
  - desinenze verbali (247)
  - lunghezza delle parole (“Mendenhall features”) (23)
- Le rappresentazioni vettoriali hanno 7,024 *feature*

# Error Rate

- L'**error rate** è una stima dell'efficacia del classificatore (0:ottima; 1:pessima) calcolata sui testi di paternità certa del *dataset*
- Nel calcolo dell'*error rate*
  - si finge di ignorare l'autore di un documento di paternità certa, si chiede al classificatore di determinarlo (usando tutti gli altri testi come *training examples*), e si controlla se la predizione del classificatore è corretta o errata
  - si ripete quanto sopra, uno a uno, per tutti i documenti del *dataset*

	<b>Error rate</b>
Classificatore per l'EPISTOLA I	0.045
Classificatore per l'EPISTOLA II	0.131

# Risultati preliminari

- **Ipotesi** indica se il classificatore ritiene che la Epistola sia da attribuirsi a Dante o NonDante
- **Confidenza** misura il “livello di convinzione” che il classificatore ha nella propria ipotesi
  - vicino a 0 o 1: alta confidenza
  - vicino a 0.5: bassa confidenza

	Ipotesi	Confidenza
CLASSIFICATORE PER L'EPISTOLA I		
CLASSIFICATORE PER L'EPISTOLA II		

# Risultati preliminari

- **Ipotesi** indica se il classificatore ritiene che la Epistola sia da attribuirsi a Dante o NonDante
- **Confidenza** misura il “livello di convinzione” che il classificatore ha nella propria ipotesi
  - vicino a 0 o 1: alta confidenza
  - vicino a 0.5: bassa confidenza

	<b>Ipotesi</b>	<b>Confidenza</b>
CLASSIFICATORE PER L'EPISTOLA I	NonDante	Bassa (0.377)
CLASSIFICATORE PER L'EPISTOLA II		

# Risultati preliminari

- **Ipotesi** indica se il classificatore ritiene che la Epistola sia da attribuirsi a Dante o NonDante
- **Confidenza** misura il “livello di convinzione” che il classificatore ha nella propria ipotesi
  - vicino a 0 o 1: alta confidenza
  - vicino a 0.5: bassa confidenza

	<b>Ipotesi</b>	<b>Confidenza</b>
CLASSIFICATORE PER L'EPISTOLA I	NonDante	Bassa (0.377)
CLASSIFICATORE PER L'EPISTOLA II	NonDante	Bassa (0.346)

# Visualizzazione (preliminare)

## Epistola I

Magnifico atque victorioso domino, domino Cani Grandi de la Scala, sacratissimi cesarei principatus in urbe Verona et civitate Vicentie vicario generali, devotissimus suus Dantes Alagherii, Florentinus natione non moribus, vitam orat per tempora diuturna felicem, et gloriosi nominis perpetuum incrementum. Inclita vestre magnificentie laus, quam fama vigil volitando disseminat, sic distrahit in diversa diversos, ut hos in spem sue prosperitatis attollat, hos extermini deiciat in terrorem. Huius quidem preconium, facta modernorum exsuperans, tanquam veri existentia latius, arbitrabar aliquando superfluum. Verum, ne diuturna me nimis incertitudo suspenderet, velut Austri regina Ierusalem petiit, velut Pallas petiit Elicona, Veronam petii fidis oculis discussurus audita, ibique magnalia vestra vidi, vidi beneficia simul et tetigi; et quemadmodum prius dictorum ex parte suspicabar excessum, sic posterius ipsa facta excessiva cognovi. Quo factum est ut ex auditu solo cum quadam animi subiectione benivolus prius exstiterim; sed ex visu postmodum devotissimus et amicus. Nec reor amicum nomen assumens, ut nonnulli forsitan obiecerent, reatum presumptionis incurrere, cum non minus dispares connectantur quam pares amicitie sacramento. Nam si delectabiles et utiles amicitias inspicere libeat, illis persepius inspicienti patebit, preheminentes inferioribus coniugari personas. Et si ad veram ac per se amicitiam torqueatur intuitus, nonne illustrium summorumque principum plerunque viros fortuna obscuros, honestate preclaros, amicos fuisse constabit? Quidni, cum etiam Dei et hominis amicitia nequaquam impediatur excessu? Quod si cuiquam, quod asseritur, nunc videretur indignum, Spiritum Sanctum audiat, amicitie sue participes quosdam homines profitentem. Nam in Sapientia de sapientia legitur, quoniam . Sed habet imperitia vulgi sine discretione iudicium; et quemadmodum solem pedalis magnitudinis arbitratur, sic et circa mores vana credulitate decipitur. Nos autem, quibus optimum quod est in nobis noscere datum est, gregum vestigia sectari non decet, quin ymo suis erroribus obviare tenemur. Nam intellectu ac ratione degentes, divina quadam libertate dotati, nullis consuetudinibus astringuntur; nec mirum, cum non ipsi legibus, sed ipsis leges potius dirigantur. Liqueat igitur, quod superius dixi, me scilicet esse devotissimum et amicum, nullatenus esse presumptum. Preferens ergo amicitiam vestram quasi thesaurum carissimum, providentia diligenti et accurata sollicitudine illam servare desidero. Itaque, cum in dogmatibus moralis negotii amicitiam adequari et salvari analogo doceatur, ad retribuendum pro collatis beneficiis plus quam semel analogam sequi michi votivum est; et propter hoc munuscula mea sepe multum consexi et ab invicem segregavi, nec non segregata percensui, dignius gratiusque vobis inquirens. Neque ipsi preheminentie vestre congruum magis comperi magis quam Comedie sublimem canticam, que decoratur titulo Paradisi; et illam sub presenti epistola, tanquam sub epigrammate proprio dedicatam, vobis ascribo, vobis offero, vobis denique recommodo. Illud quoque preterire silentio simpliciter inardescens non sinit affectus, quod in hac donatione plus dono quam domino et honoris et fame conferri potest videri. Quidni cum eius titulum iam presagiam de gloria vestri nominis ampliandum? Satis actenus videbar expressisse quod de proposito fuit; sed zelus gratie vestre, quam sitio quasi vitam parvipendens, a primordio metam prefixam urget ulterius. Itaque, formula consumata epistole, ad introductionem oblatis operis aliquid sub lectoris officio compendiose aggrediar. In parte vero executiva, que fuit divisa contra totum prologum, nec dividendo nec sententiando quicquam dicitur ad presens, nisi hoc, quod ubique proceditur ascendendo de celo in celum, et reccatur de animabus beatis inventis in quolibet orbe. Et quia illa vera beatitudo in sentiendo veritatis principio consistit - ut patet per Iohannem ibi: . et cetera; et per Boetium in tertio De Consolatione ibi: -, inde est quod, ad ostendendum gloriam beatitudinis in illis animabus, ab eis tanquam videntibus omnem veritatem multa queruntur, que magnam habent utilitatem et delectationem. Et quia, invento principio seu primo, videlicet Deo, nichil est quod ulterius queratur, cum sit Alpha et O, id est principium et finis, ut visio Iohannis designat, in ipso Deo terminatur tractatus, qui est benedictus in secula seculorum.





# Osservazioni conclusive

Differenze macroscopiche fra le ricerche sull'attribuzione sviluppate secondo i metodi filologici (in senso lato) e le ricerche sull'attribuzione fondate sul *machine learning*. Si tratta di procedure di ricerca radicalmente diverse.

- Ricerche filologiche (in senso lato):
  - Argomenti interni ai testi:
    - Se una certa idea o affermazione del testo è coerente o compatibile con quanto sappiamo del pensiero dell'autore
    - Se una certa espressione del testo è coerente o compatibile con la lingua e lo stile dell'autore
  - Argomenti interni/esterni ai testi:
    - Se questo o quel dato del testo, o il testo nel suo insieme, sono compatibili con le circostanze biografiche nelle quali l'autore potrebbe aver composto il testo
    - Se la tradizione del testo è compatibile con le circostanze biografiche nelle quali l'autore potrebbe aver composto il testo
  - Ecc.

# Osservazioni conclusive

- *Authorship analysis*:
  - Si basa esclusivamente sul confronto tra caratteristiche oggettive del testo in questione e caratteristiche oggettive dei testi certi dell'autore o degli autori candidati.
  - Cioè si basa su caratteristiche di livello basico ed estese sull'intero testo (sequenza di caratteri)
  - Non è escluso che possa basarsi anche su caratteristiche di livello più alto, frutto di una elaborazione, purché sempre oggettive ed estese sull'intero testo, ma la pratica corrente predilige le caratteristiche basiche, materiali.

# Osservazioni conclusive

Dunque:

- Ricerche filologiche (in senso lato):
  - Operano su argomenti singoli, puntuali, di varia natura
  - Elaborati da una intelligenza umana su vari livelli di astrazione e di complessità in base a criteri logici espliciti o esplicitabili così da tendere a risultati “falsificabili”
- *Authorship analysis*:
  - Opera su caratteristiche oggettive dei testi, estese sugli interi testi, preferibilmente caratteristiche materiali di livello basico (sequenze di caratteri)
  - Il confronto viene sviluppato dalla macchina in base ad algoritmi che costituiscono una sua “soggettività” (si sarà notato che il grado di confidenza che la macchina attribuisce a ogni propria decisione è una probabilità “soggettiva”)

# Osservazioni conclusive

Una prima conclusione:

- Data questa radicale diversità, le procedure e i risultati della *authorship analysis* si collocano su un canale parallelo e – almeno inizialmente – non comunicante con le procedure e i risultati della discussione scientifica fondata su argomenti filologici (in senso lato).
- In altre parole, le decisioni a cui perviene la *authorship analysis* NON avvalorano né inficiano la validità di alcun singolo argomento addotto pro o contro la paternità dantesca.

## Osservazioni conclusive

Che validità dobbiamo riconoscere alle decisioni a cui perviene la *authorship analysis*?

- Se guardiamo alle caratteristiche su cui essa opera (caratteristiche oggettive, materiali, basiche, in pratica sequenze di caratteri) è più che legittimo il dubbio se e quanto esse racchiudano “l'impronta digitale” dell'autore
- Se guardiamo alla logica “soggettiva” con cui la macchina trae le sue conclusioni dalle caratteristiche di cui sopra, logica che non siamo sempre in grado di controllare fino in fondo, è più che legittimo il dubbio se e quanto essa sia affidabile.

## Osservazioni conclusive

La risposta a entrambi questi dubbi è che, prima di applicare i suoi strumenti di decisione al caso dubbio, la macchina li ha applicati a casi non dubbi (casi noti per noi, ma non per la macchina), e i risultati a cui è pervenuta sono corretti. Cioè

- dopo aver sottratto ai testi di Dante una certa porzione di testo, è stato chiesto alla macchina se quella porzione di testo sia di Dante o NonDante, e la macchina ha risposto che è di Dante
- dopo aver sottratto ai testi di NonDante una certa porzione di testo, è stato chiesto alla macchina se quella porzione di testo sia di Dante o NonDante, e la macchina ha risposto che è di NonDante. Quindi la macchina si è guadagnata una non trascurabile credibilità.

# Osservazioni conclusive

- Ciò premesso, i risultati di questo lavoro sono preliminari e devono essere interpretati come tali
- Limiti attuali della ricerca:
  - Limitato numero di test di confronto
  - Limitato numero di tipi di *feature*
  - Basso grado di confidenza dei risultati



# Osservazioni conclusive

Cosa intendiamo fare per arricchire e consolidare la ricerca e giungere a risultati più sicuri e significativi nel prossimo futuro:

- Aumentare i testi di confronto, costituendo *dataset* più ampi e più rappresentativi:
  - della epistolografia tra fine Duecento e Trecento (da confrontare con Epistola I)
  - degli *accessus* e di altre scritture esegetiche e retoriche tra fine Duecento e Trecento (da confrontare con Epistola II)
- Aumentare le *feature* includendo elementi linguistici di secondo livello, non immediatamente coincidenti con sequenze di caratteri, estraibili dai testi grazie all'applicazione di strumenti computazionali di NLP relativi alla lingua latina, p.es.:
  - suffissi (nel senso di morfemi, non di sequenze materiali di caratteri)
  - *POS tags*
  - indici di frequenza delle parole
- Chiedere al sistema, via via che riusciremo a innalzare il grado di confidenza dei risultati, quali *feature* risultino più rivelatrici
- Esplorare la bibliografia sul *machine learning* cercando di individuare strade nuove per superare lo stato dell'arte.

Grazie per l'attenzione