

*BLENDING AUTHORSHIP ATTRIBUTION.  
UNMASKING ELENA FERRANTE  
COMBINING DIFFERENT AUTHOR  
PROFILING METHODS*

George K. Mikros

Department of Italian Language and  
Literature

National and Kapodistrian University of  
Athens

Greece

# ELENA FERRANTE'S CORPUS

## Basic stats

- 40 authors
- 150 novels
- Total 9,662,611 words
- Texts of variable length (Min= 8,129 words, Max= 194,993 words, St.Dev= 38,366)
- Ferrante is represented by 7 novels (623,466 words)
- Ferrante's texts are also variable in size (Min= 36,091 words, Max= 139,491 words, St.Dev.=45,282)

# THE AUTHORSHIP PROBLEM

The Ferrante's problem as it has been posed could be viewed as a closed classification problem with the premise that...

**we are certain** that one of the 39 authors is the one behind Ferrante

However...

Is that the case?

Can we develop a methodology that will be generic enough so that it can handle this problem as closed-classification even when we are not sure if the real author is among the candidates?

# STANDARD AUTHORSHIP ATTRIBUTION METHOD

## Standard Authorship Attribution method:

- Train an algorithm (distance based, machine learning etc.) in a dataset where each text has a known author and the text is represented by a vector of measurements in various stylometric features.
- The algorithm “learns” the relation of the author with its quantitative linguistic profile.
- The “expert” algorithm tests its “knowledge” to a new quantitative linguistic profile and determines which author (existing in his training set) is the most probable.

|    | A   | B        | C           | D           | E            | F           | G            | H           | I         | J         | K           | L           | M           |
|----|---|----------|-------------|-------------|--------------|-------------|--------------|-------------|-----------|-----------|-------------|-------------|-------------|
|    | Filename                                  | Author   | wug1        | wug2        | wug3         | wug4        | wug5         | wug6        | wug7      | wug8      | wug9        | wug10       | wug11       |
| 1  | Affinati_1997_Campo_del_sangue.txt_0.txt  | Affinati | 4.4         | 1.4         | 1.7          | 2           | 1.6          | 1.8         | 1.4       | 0.7       | 1.3         | 0.5         | 0.4         |
| 2  | Affinati_1997_Campo_del_sangue.txt_1.txt  | Affinati | 3.7         | 1.8         | 1.2          | 2.3         | 1.8          | 1.9         | 1.1       | 0.7       | 1.8         | 0.8         | 0.1         |
| 3  | Affinati_1997_Campo_del_sangue.txt_10.txt | Affinati | 3.5         | 1.7         | 1.6          | 1.8         | 1.7          | 1.9         | 1.5       | 0.7       | 2.3         | 0.3         | 0.1         |
| 4  | Affinati_1997_Campo_del_sangue.txt_11.txt | Affinati | 3.5         | 1.4         | 1            | 1.7         | 2.2          | 3.4         | 1.5       | 0.3       | 2.2         | 0.5         | 0.8         |
| 5  | Affinati_1997_Campo_del_sangue.txt_12.txt | Affinati | 3.2         | 1           | 2.5          | 1.2         | 1.4          | 2.1         | 1.5       | 0.4       | 2           | 0.9         | 1           |
| 6  | Affinati_1997_Campo_del_sangue.txt_13.txt | Affinati | 4.7         | 1.8         | 2            | 2.6         | 1.7          | 2           | 1.1       | 0.6       | 1.4         | 1.2         | 0.6         |
| 7  | Affinati_1997_Campo_del_sangue.txt_14.txt | Affinati | 3.3         | 1.9         | 1.9          | 2.2         | 0.8          | 1.3         | 0.9       | 0.7       | 2.2         | 0.4         | 0.1         |
| 8  | Affinati_1997_Campo_del_sangue.txt_15.txt | Affinati | 3.1         | 0.8         | 1.2          | 1.3         | 1.9          | 2.5         | 1.1       | 0.6       | 1           | 0.9         | 0.1         |
| 9  | Affinati_1997_Campo_del_sangue.txt_16.txt | Affinati | 3.2         | 0.9         | 1.6          | 2.2         | 2            | 1.6         | 1.1       | 0.8       | 1.9         | 0.7         | 0.4         |
| 10 | Affinati_1997_Campo_del_sangue.txt_17.txt | Affinati | 3.4         | 1.3         | 2            | 2.4         | 1.5          | 1.2         | 1.3       | 1         | 1.9         | 1           | 0.6         |
| 11 | Affinati_1997_Campo_del_sangue.txt_18.txt | Affinati | 3.5         | 1.8         | 2.1          | 1.7         | 1.6          | 1.9         | 1.4       | 0.7       | 1.9         | 0.8         | 1.1         |
| 12 | Affinati_1997_Campo_del_sangue.txt_19.txt | Affinati | 2.3         | 2           | 1.4          | 3.2         | 1.2          | 2           | 2         | 0.6       | 2.5         | 0.6         | 0.1         |
| 13 | Affinati_1997_Campo_del_sangue.txt_2.txt  | Affinati | 2.8         | 2.3         | 1.7          | 1.9         | 0.8          | 1.8         | 1.3       | 0.4       | 2.1         | 0.8         | 1.4         |
| 14 | Affinati_1997_Campo_del_sangue.txt_20.txt | Affinati | 2.4         | 0.9         | 1.4          | 2.5         | 1.7          | 1.9         | 1.6       | 1.1       | 1.8         | 0.9         | 1.1         |
| 15 | Affinati_1997_Campo_del_sangue.txt_21.txt | Affinati | 3.9         | 1           | 1.9          | 2.5         | 2.7          | 2.2         | 1.1       | 1.1       | 1.3         | 1           | 0.1         |
| 16 | Affinati_1997_Campo_del_sangue.txt_22.txt | Affinati | 3.8         | 1.9         | 2            | 2.4         | 1.8          | 2.1         | 1.1       | 1.1       | 2.6         | 0.5         | 0.1         |
| 17 | Affinati_1997_Campo_del_sangue.txt_23.txt | Affinati | 3.9         | 2           | 1.8          | 2.4         | 1.4          | 3.3         | 1.9       | 0.6       | 2.3         | 0.5         | 1           |
| 18 | Affinati_1997_Campo_del_sangue.txt_24.txt | Affinati | 3.2         | 0.9         | 1.5          | 1.9         | 2.4          | 3.2         | 1         | 0.5       | 1.9         | 0.4         | 2.1         |
| 19 | Affinati_1997_Campo_del_sangue.txt_25.txt | Affinati | 3.2         | 2           | 1.1          | 2.4         | 2.3          | 1.4         | 0.8       | 1         | 2.5         | 0.5         | 0.8         |
| 20 | Affinati_1997_Campo_del_sangue.txt_26.txt | Affinati | 3.8         | 1.3         | 1.3          | 1.3         | 1.7          | 1.9         | 1.1       | 0.5       | 2.4         | 0.7         | 0.8         |
| 21 | Affinati_1997_Campo_del_sangue.txt_27.txt | Affinati | 4.1         | 1.6         | 1.6          | 1.6         | 1.2          | 1.5         | 1.4       | 0.8       | 1.7         | 0.5         | 0.8         |
| 22 | Affinati_1997_Campo_del_sangue.txt_28.txt | Affinati | 3           | 1.2         | 0.9          | 2           | 1.4          | 2.2         | 1.2       | 0.7       | 2.8         | 0.6         | 1.1         |
| 23 | Affinati_1997_Campo_del_sangue.txt_29.txt | Affinati | 3.7         | 2           | 1.2          | 1.6         | 1.1          | 2           | 0.8       | 0.8       | 1.7         | 0.9         | 1.1         |
| 24 | Affinati_1997_Campo_del_sangue.txt_3.txt  | Affinati | 3.8         | 1.3         | 1.4          | 2.1         | 1.6          | 1.4         | 1.2       | 0.6       | 1.4         | 0.6         | 0.1         |
| 25 | Affinati_1997_Campo_del_sangue.txt_20.txt | Affinati | 3.5         | 1.8         | 1.4          | 2.4         | 1.5          | 1.5         | 1.2       | 0.5       | 1           | 0.9         | 0.1         |
| 26 | Affinati_1997_Campo_del_sangue.txt_31.txt | Affinati | 3.3         | 2           | 1            | 1.6         | 1.6          | 1.9         | 1.4       | 0.8       | 2           | 0.5         | 0.1         |
| 27 | Affinati_1997_Campo_del_sangue.txt_32.txt | Affinati | 3.3         | 1.6         | 1.6          | 2.4         | 1.5          | 2           | 1.4       | 0.9       | 1.6         | 1           | 0.1         |
| 28 | Affinati_1997_Campo_del_sangue.txt_33.txt | Affinati | 2.4         | 0.9         | 1.4          | 2           | 1.6          | 2.6         | 1.4       | 0.7       | 2           | 0.7         | 0.1         |
| 29 | Affinati_1997_Campo_del_sangue.txt_34.txt | Affinati | 3.9         | 1.8         | 2.1          | 3.1         | 1.6          | 1.7         | 0.8       | 0.9       | 1.9         | 0.6         | 0.1         |
| 30 | Affinati_1997_Campo_del_sangue.txt_35.txt | Affinati | 3.696303696 | 1.298701299 | 1.1988801199 | 1.798201798 | 1.1988801199 | 1.498501499 | 0.2997003 | 0.4995005 | 1.498501499 | 0.699300699 | 0.899100899 |
| 31 | Affinati_1997_Campo_del_sangue.txt_36.txt | Affinati | 4.3         | 0.8         | 1.8          | 2.3         | 1.3          | 1.5         | 1.2       | 1.2       | 2.1         | 1.1         | 0.1         |
| 32 | Affinati_1997_Campo_del_sangue.txt_37.txt | Affinati | 3.7         | 1.8         | 1.4          | 1.8         | 2.2          | 1.7         | 0.5       | 0.7       | 1.7         | 0.6         | 0.1         |
| 33 | Affinati_1997_Campo_del_sangue.txt_4.txt  | Affinati | 3.3         | 1.7         | 1.6          | 2.3         | 2            | 2.1         | 1.1       | 0.9       | 1.8         | 0.8         | 0.4         |
| 34 | Affinati_1997_Campo_del_sangue.txt_5.txt  | Affinati | 4.3         | 1.5         | 1.5          | 2.9         | 1.8          | 2           | 1.2       | 1.1       | 1.4         | 1.1         | 1.1         |
| 35 | Affinati_1997_Campo_del_sangue.txt_6.txt  | Affinati | 2.7         | 1.3         | 1.8          | 2.2         | 2.5          | 2.7         | 0.9       | 1.2       | 1.6         | 0.5         | 0.1         |
| 36 | Affinati_1997_Campo_del_sangue.txt_7.txt  | Affinati | 3.4         | 1.1         | 1.7          | 1.6         | 3            | 2.2         | 1.5       | 1         | 2           | 0.2         | 0.6         |

# BLENDED AUTHORSHIP ATTRIBUTION METHOD

Combine multiple Author Profiling Classifications in order to reduce the number of candidate authors.

Since the Author Profiling is based on more generic categories (Author's Gender, Age, Region etc) we can use it in a more generic way and in cases where the author is not inside the candidates corpus.

For example: An author that has been profiled as Female, under 40 and has been born in the region of Lazio could be identified as the Valerio in our dataset. But, we could not have an author with this profile in our corpus. Thus this method has developed a frame of author characteristics that can be used to test authorship and identify authors even when they don't belong to our dataset but we have theoretical reasons to compare them.

# THE BLENDED METHOD IN ACTION

We performed author profiling classifications in our datasets as follows:

- Gender profiling: Learn our Machine Learning Algorithm (MLA) to recognize the author's gender.
- Age profiling: Learn our MLA to identify author's age group.
- Region profiling: Learn our MLA to identify the author's region.
- Author's town profiling: Learn our MLA to identify the author's town .

The developed profiling models were used to predict Elena Ferrante's

- Gender
- Age
- Region
- Town

The intersection of this characteristics gives us the author behind Ferrante's pseudonym.

# THE TECHNICAL STUFF...

## Corpus preprocessing

- Tokenization using regexes optimized for the Italian language
- Each novel was sliced in 1,000 words chunks so that our vector sample gets bigger (150 => 9,514)

## Stylometric Features (Total 4,000 features)

- Most frequent words (1,000)
- Author Multilevel N-gram profiles (AMNP)
  - Character 2-grams (1,000)
  - Character 3-grams (1,000)
  - Word 2-grams (1,000)

# MORE TECHNICAL STUFF....

## Machine Learning

- MLA: Support Vector Machines (SVM) using Polynomial Kernel
- Scaling: All data were scaled to z-values so that all values were normalized in the z scale.
- Feature reduction: Due to data sparsity we used near-zero variance predictor detection (4,000 => 2,672).
  - The percentage of unique values is less than 20% and
  - The ratio of the most frequent to the second most frequent value is greater than 20
- MLA Evaluation: The SVM performance was evaluated using 10-fold cross-validation.

## Software

- All experiments were conducted using the language R and the R Studio IDE
- The MLA and the predictions were performed using the *caret* package

# GENDER PROFILING

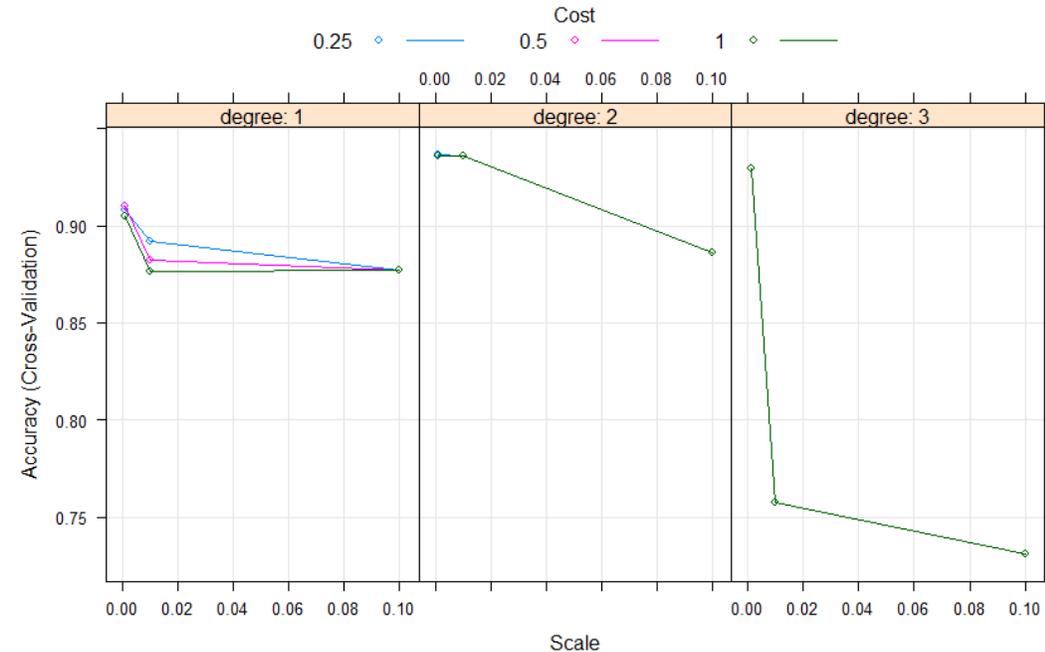
We trained our *MLA* to identify Author's gender using our training corpus.

Multiple SVM models were developed so that the best SVM parameters could be combined.

The cross-validation accuracy of the optimized model was: **0.94**

**Ferrante: Male**

- **594 of 619 chunks (96%)**



# GENDER RESTRICTION

## Male authors

|            |           |          |
|------------|-----------|----------|
| Affinati   | DeLuca    | Piccolo  |
| Ammaniti   | DeSilva   | Pincio   |
| Bajani     | Faletti   | Prisco   |
| Balzano    | Fois      | Raimo    |
| Baricco    | Giordano  | ReaE     |
| Benni      | Lagioia   | Scarpa   |
| Brizzi     | Montesano | Starnone |
| Carofiglio | Nesi      | Vasta    |
| Covacich   | Nori      | Veronesi |

## Female authors

|            |         |
|------------|---------|
| Maraini    | Sereni  |
| Mazzantini | Tamaro  |
| Mazzucco   | Valerio |
| Milone     | Vinci   |
| Morazzoni  |         |
| Murgia     |         |
| Parrella   |         |
| Ramondino  |         |

# GENDER RESTRICTION

## Male authors

|            |           |          |
|------------|-----------|----------|
| Affinati   | DeLuca    | Piccolo  |
| Ammaniti   | DeSilva   | Pincio   |
| Bajani     | Faletti   | Prisco   |
| Balzano    | Fois      | Raimo    |
| Baricco    | Giordano  | ReaE     |
| Benni      | Lagioia   | Scarpa   |
| Brizzi     | Montesano | Starnone |
| Carofiglio | Nesi      | Vasta    |
| Covacich   | Nori      | Veronesi |

## Female authors

|            |         |
|------------|---------|
| Maraini    | Sereni  |
| Mazzantini | Tamaro  |
| Mazzucco   | Valerio |
| Milone     | Vinci   |
| Morazzoni  |         |
| Murgia     |         |
| Parrella   |         |
| Ramondino  |         |

# REGION PROFILING

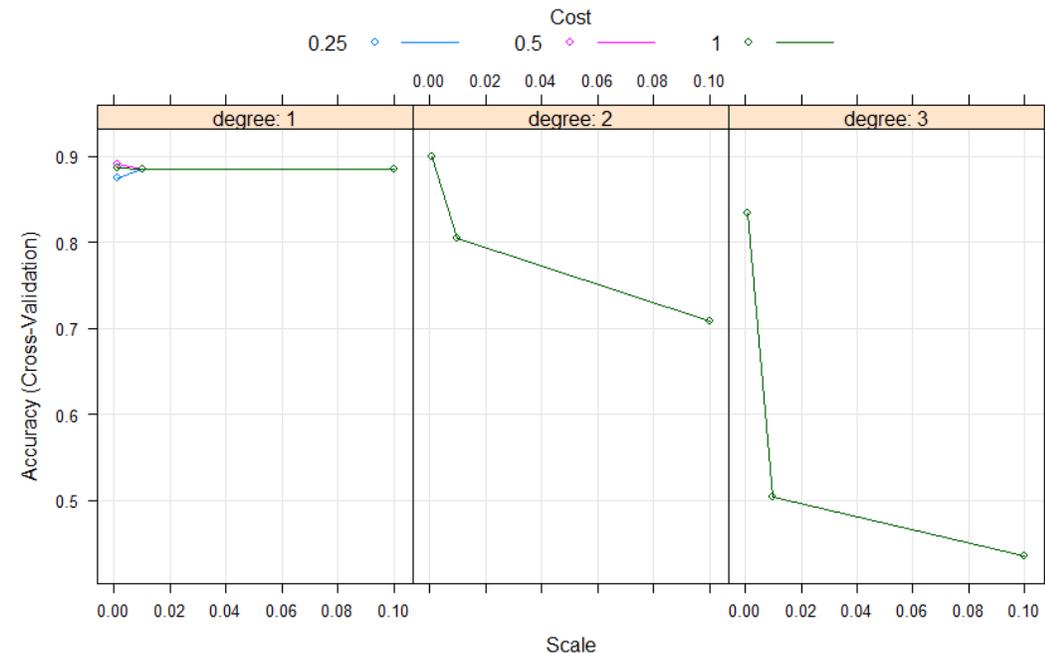
We trained our *MLA* to identify Author's region using our training corpus.

Multiple SVM models were developed so that the best SVM parameters could be combined.

The cross-validation accuracy of the optimized model was: **0.90**

**Ferrante: Campania**

- **607 of 619 chunks (98%)**



# GENDER - REGION RESTRICTION

## Male authors from Campania

DeLuca

DeSilva

Montesano

Piccolo

Prisco

ReaE

Starnone

## Male authors from other parts of Italy

Affinati

Covacich

Pincio

Ammaniti

Faletti

Raimo

Bajani

Ferrante

Scarpa

Balzano

Fois

Vasta

Baricco

Giordano

Veronesi

Benni

Lagioia

Brizzi

Nesi

Carofiglio

Nori

# GENDER - REGION RESTRICTION

## Male authors from Campania

DeLuca

DeSilva

Montesano

Piccolo

Prisco

ReaE

Starnone

## Male authors from other parts of Italy

Affinati

Covacich

Pincio

Ammaniti

Faletti

Raimo

Bajani

Ferrante

Scarpa

Balzano

Fois

Vasta

Baricco

Giordano

Veronesi

Benni

Lagioia

Brizzi

Nesi

Carofiglio

Nori

# TOWN PROFILING

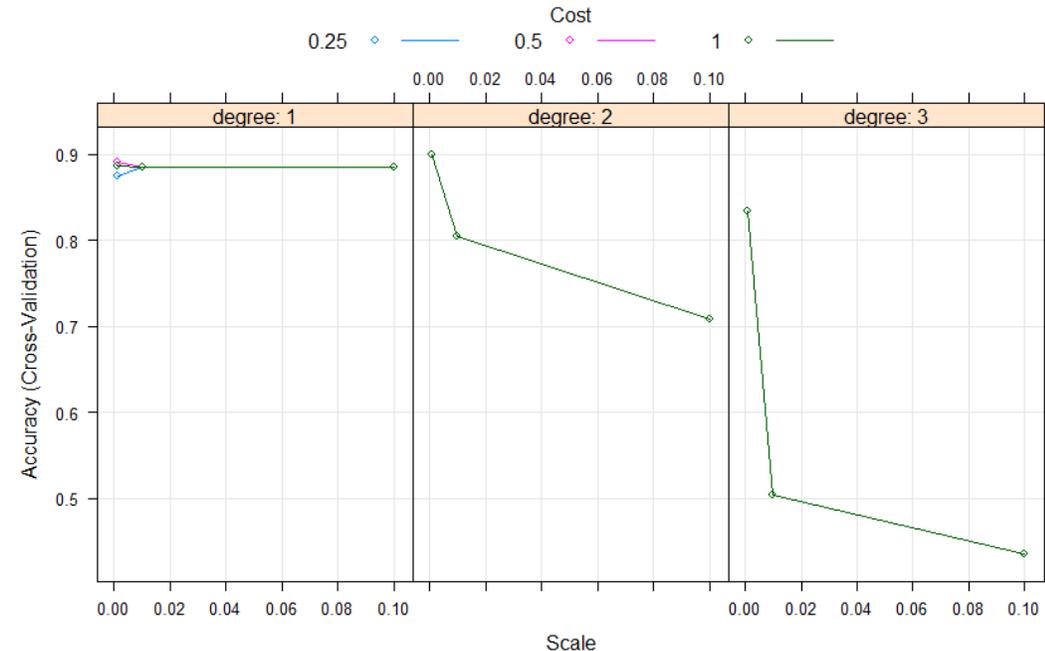
We trained our *MLA* to identify Author's town using our training corpus.

Multiple SVM models were developed so that the best SVM parameters could be combined.

The cross-validation accuracy of the optimized model was: **0.94**

**Ferrante: Saviano = Starnone**

- **608 of 619 chunks (98%)**



# GENDER — REGION > TOWN RESTRICTION

## Male authors from Saviano

Starnone

## Male authors from other parts of Italy

|            |           |          |
|------------|-----------|----------|
| Affinati   | DeLuca    | Piccolo  |
| Ammaniti   | DeSilva   | Pincio   |
| Bajani     | Faletti   | Prisco   |
| Balzano    | Fois      | Raimo    |
| Baricco    | Giordano  | ReaE     |
| Benni      | Lagioia   | Scarpa   |
| Brizzi     | Montesano | Vasta    |
| Carofiglio | Nesi      | Veronesi |
| Covacich   | Nori      |          |

# GENDER — REGION > TOWN RESTRICTION

Male authors from Saviano

**Starnone**

Male authors from other parts of Italy

|            |           |          |
|------------|-----------|----------|
| Affinati   | DeLuca    | Piccolo  |
| Ammaniti   | DeSilva   | Pincio   |
| Bajani     | Faletti   | Prisco   |
| Balzano    | Fois      | Raimo    |
| Baricco    | Giordano  | ReaE     |
| Benni      | Lagioia   | Scarpa   |
| Brizzi     | Montesano | Vasta    |
| Carofiglio | Nesi      | Veronesi |
| Covacich   | Nori      |          |

# AGE PROFILING

We trained our MLA to identify Author's age group using our training corpus.

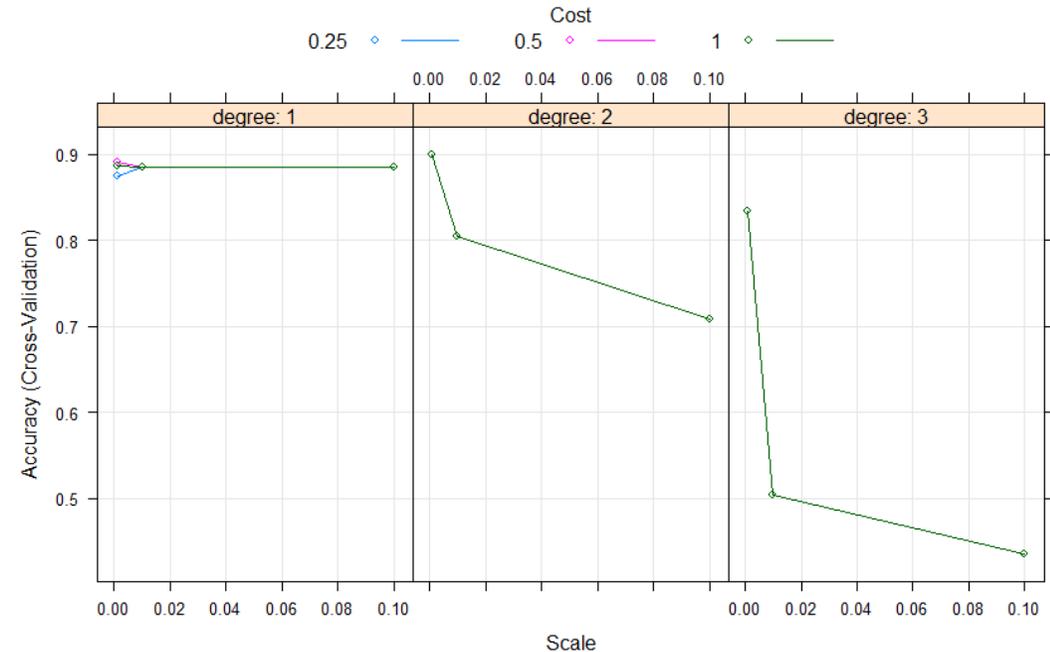
The authors of our dataset were categorized in 3 age groups: < 40, 40-60 and > 60

Multiple SVM models were developed so that the best SVM parameters could be combined.

The cross-validation accuracy of the optimized model was: **0.92**

**Ferrante: > 60**

- **561 of 619 chunks (91%)**



# GENDER — REGION — AGE RESTRICTION

## Male authors from Campania over 60

DeLuca

Prisco

ReaE

Starnone

## Male authors from other parts of Italy less than 60

Affinati

Covacich

Nori

Ammaniti

DeSilva

Piccolo

Bajani

Faletti

Pincio

Balzano

Fois

Raimo

Baricco

Giordano

Scarpa

Benni

Lagioia

Vasta

Brizzi

Montesano

Veronesi

Carofiglio

Nesi

# GENDER — REGION — AGE RESTRICTION

## Male authors from Campania over 60

DeLuca

Prisco

ReaE

Starnone

## Male authors from other parts of Italy less than 60

Affinati

Covacich

Nori

Ammaniti

DeSilva

Piccolo

Bajani

Faletti

Pincio

Balzano

Fois

Raimo

Baricco

Giordano

Scarpa

Benni

Lagioia

Vasta

Brizzi

Montesano

Veronesi

Carofiglio

Nesi

# AUTHORSHIP ATTRIBUTION IN THE FILTERED LIST OF AUTHORS

We trained our *MLA* to identify the Author using our training corpus (contains only 4 candidates).

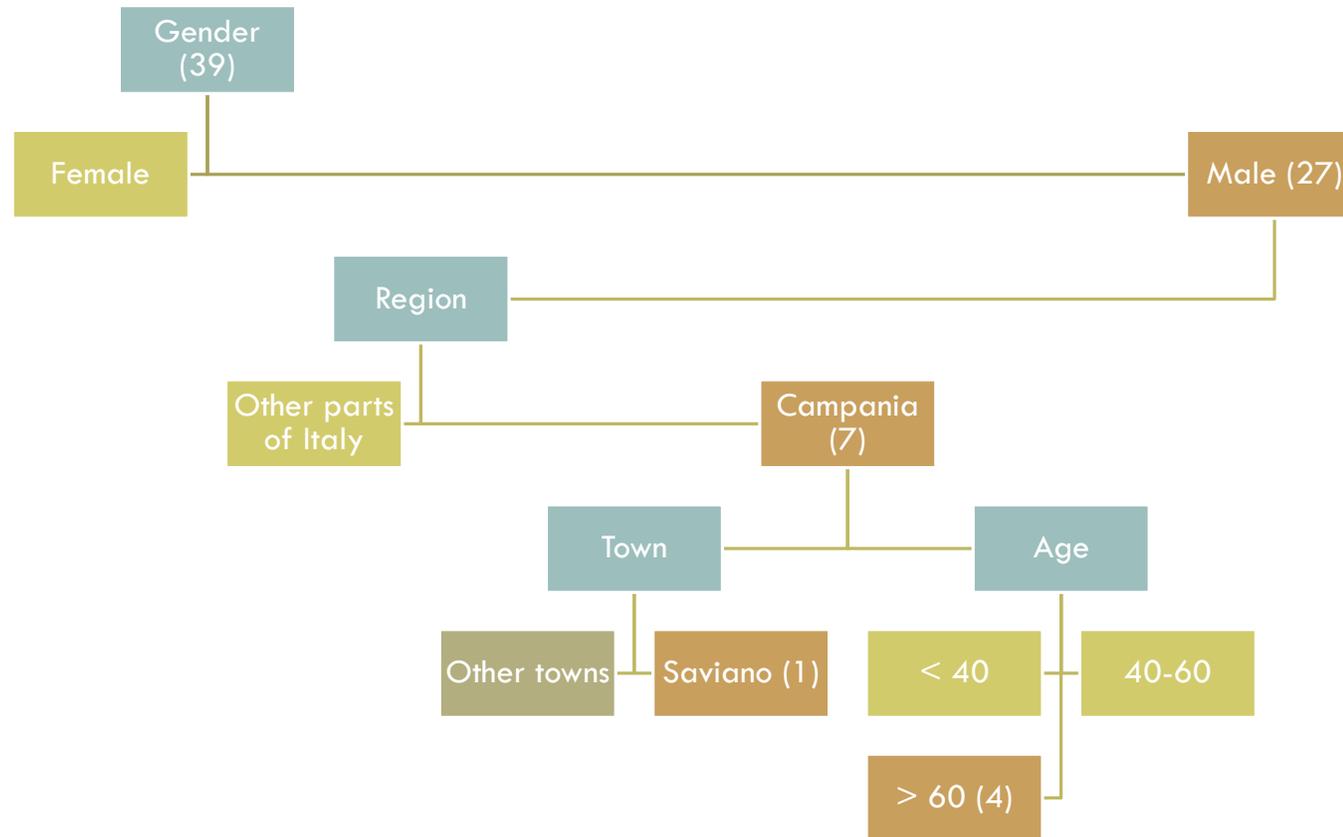
Multiple SVM models were developed so that the best SVM parameters could be combined.

The cross-validation accuracy of the optimized model was: **0.996**

**Ferrante: Starnone**

- **619 of 619 chunks (100%)**

# NAVIGATING THE PROFILING RESTRICTIONS



# CONCLUSIONS

We presented a blended authorship attribution method where multiple author profiling classification restrict the initial sample of candidate authors to a few or even the one most probable real author.

This method is generic in the sense that we develop a frame of author's characteristics that can reliably identify the real author even when he/she is not among the candidates.

Elena Ferrante's profile models were very accurate ( $> 90\%$ ) and indicate that she is:

- Male
- From Campania
- From the town Saviano
- Over 60 years old

**Elena Ferrante stylometric profile matches perfectly Domenico Starnone**

**Thank you !!!!**