AUTHOR ATTRIBUTION WITH MACHINE LEARNING

Melek Nida Yılmaz

What is Author Attribution?

- **Author’s Fingerprint**
- **Idiolect Hypothesis**
- **Syntactic and Lexical pattern**
- **Unconscious language habits**

⋆ I ate because I was hungry.

⋆ Because I was hungry, I ate.

⋆ However → Thus

Introduction to Methods

Historical Aspect

- Statistical feature extraction
- Frequency distribution
- Text → Numerical Vectors

- **TF-IDF (TermFrequency-InverseDocumentFrequency)**

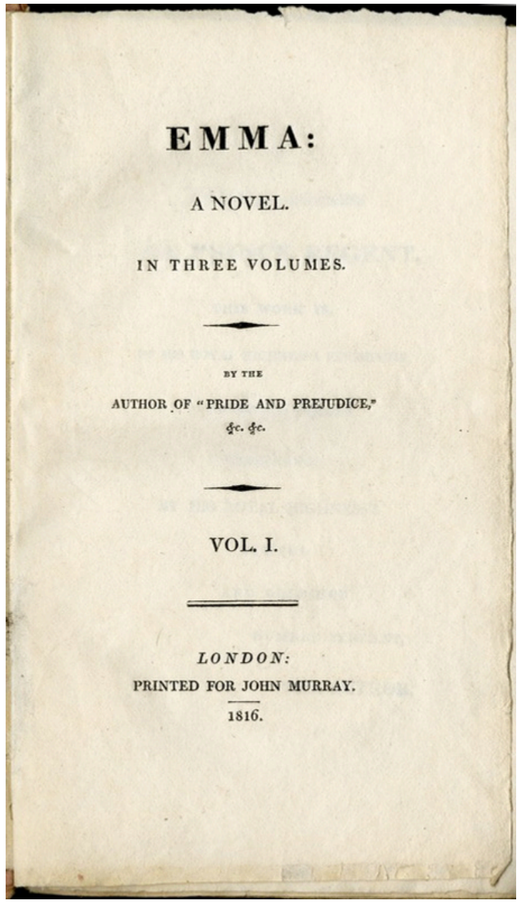
- **Burrows' Delta**

Advanced Methodologies

- Transformer
- Embedding
- Semantic Relationships

- **BERT (Bidirectional Encoder Representations from Transformers)**

What is our Case?



EMMA:

A NOVEL.

IN THREE VOLUMES.

BY THE
AUTHOR OF "PRIDE AND PREJUDICE,"
&c. &c.

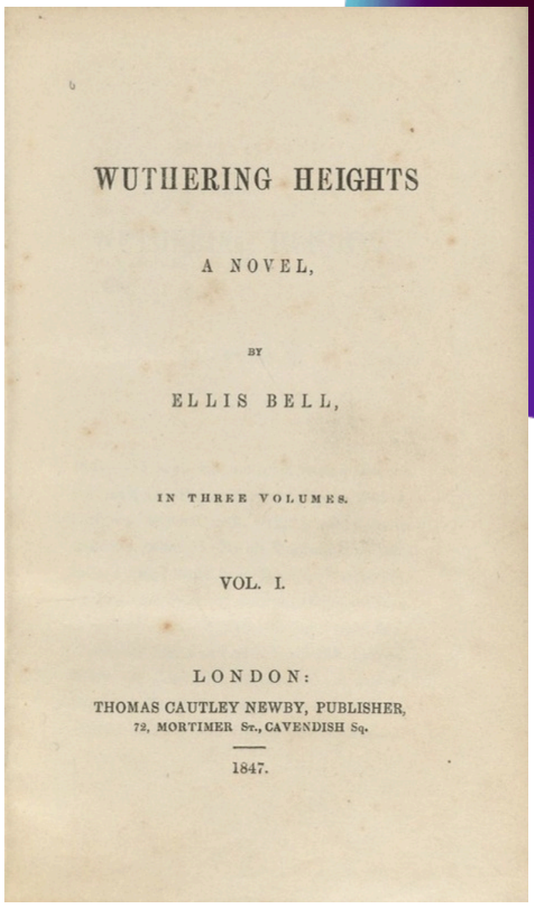
VOL. I.

LONDON:
PRINTED FOR JOHN MURRAY.
1816.

- Comparative analysis
- Accuracy
- Determine stylistic features with mathematical framework

- 1815-1847 (It)
- Countryside of England
- Female Perspective

- Pride and Prejudice - Jane Austen
- Jane Eyre - Charlotte Brontë
- Dune - Sci-fi Novel



WUTHERING HEIGHTS

A NOVEL,

BY

ELLIS BELL,

IN THREE VOLUMES.

VOL. I.

LONDON:
THOMAS CAUTLEY NEWBY, PUBLISHER,
72, MORTIMER Sr., CAVENDISH Sq.
1847.

TF-IDF and Vector Space Model

terms

Documents

$$idf(t, \mathcal{D}) = \log\left(\frac{N}{|\{d \in \mathcal{D} : t \in d\}|}\right) + 1$$

$$w_{t,d} = tf(t, d) \cdot idf(t, \mathcal{D})$$

term frequency

Burrows' Delta

$$z_i(d) = \frac{f_i(d) - \mu_i}{\sigma_i}$$

Z-normalization

The usage frequency of function words follows a Gaussian distribution

$$\Delta_{(T,A)} = \frac{1}{n} \sum_{i=1}^{n} |z_i(T) - z_i(A)|$$

Manhattan Distance

Difference between centroid and every other word

BERT

mutli-head self-attention mechanism

Query

Value

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Key

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

Embedding matrice

Weighted matrice

Fine-Tuning

- Discriminative fine-tuning
- Extra Classification Layer
- adding token: $h_{CLS} \in \mathbb{R}^H$

$$P(y \mid x) = \text{softmax}(W h_{CLS} + b)$$

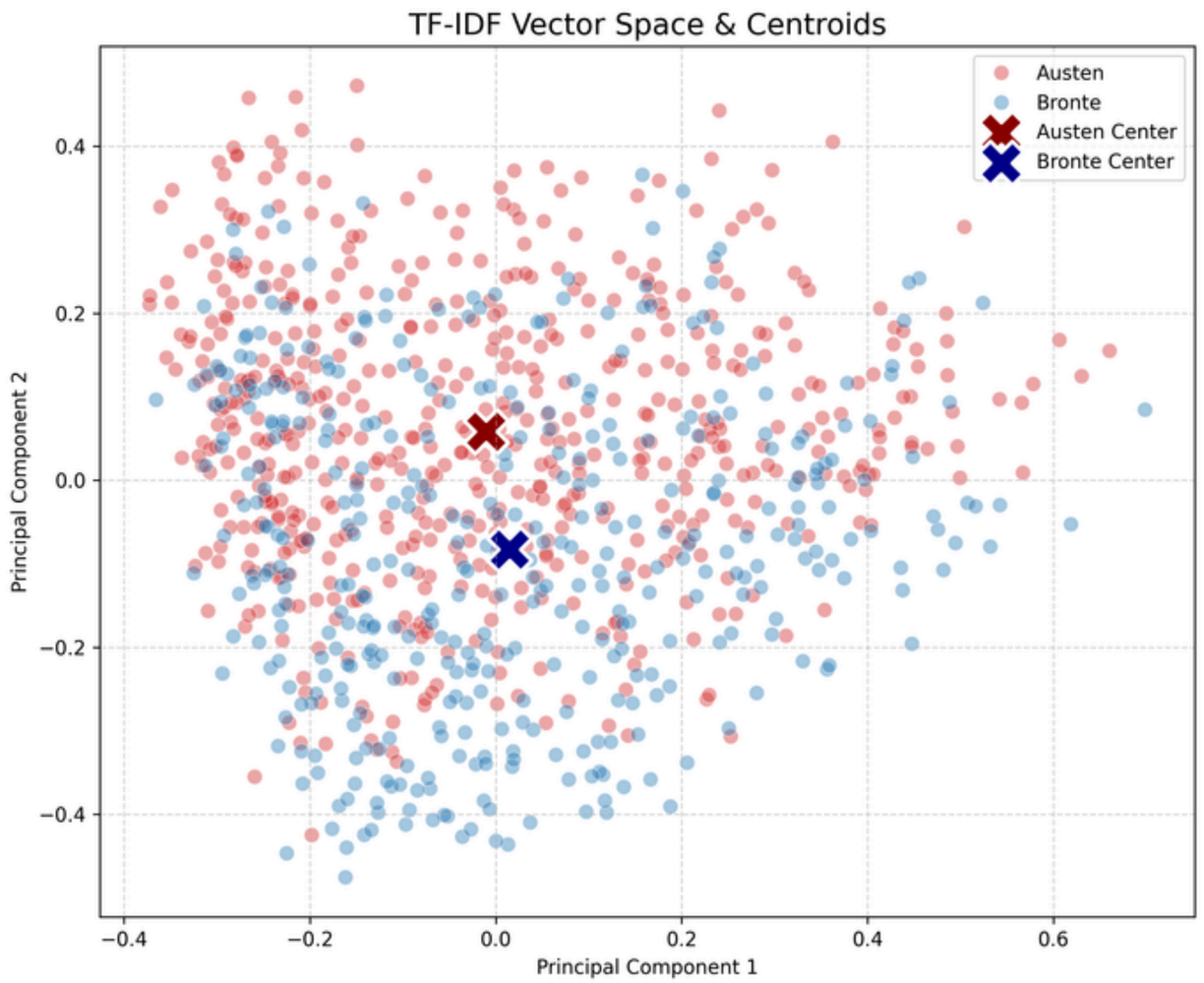
Author

input

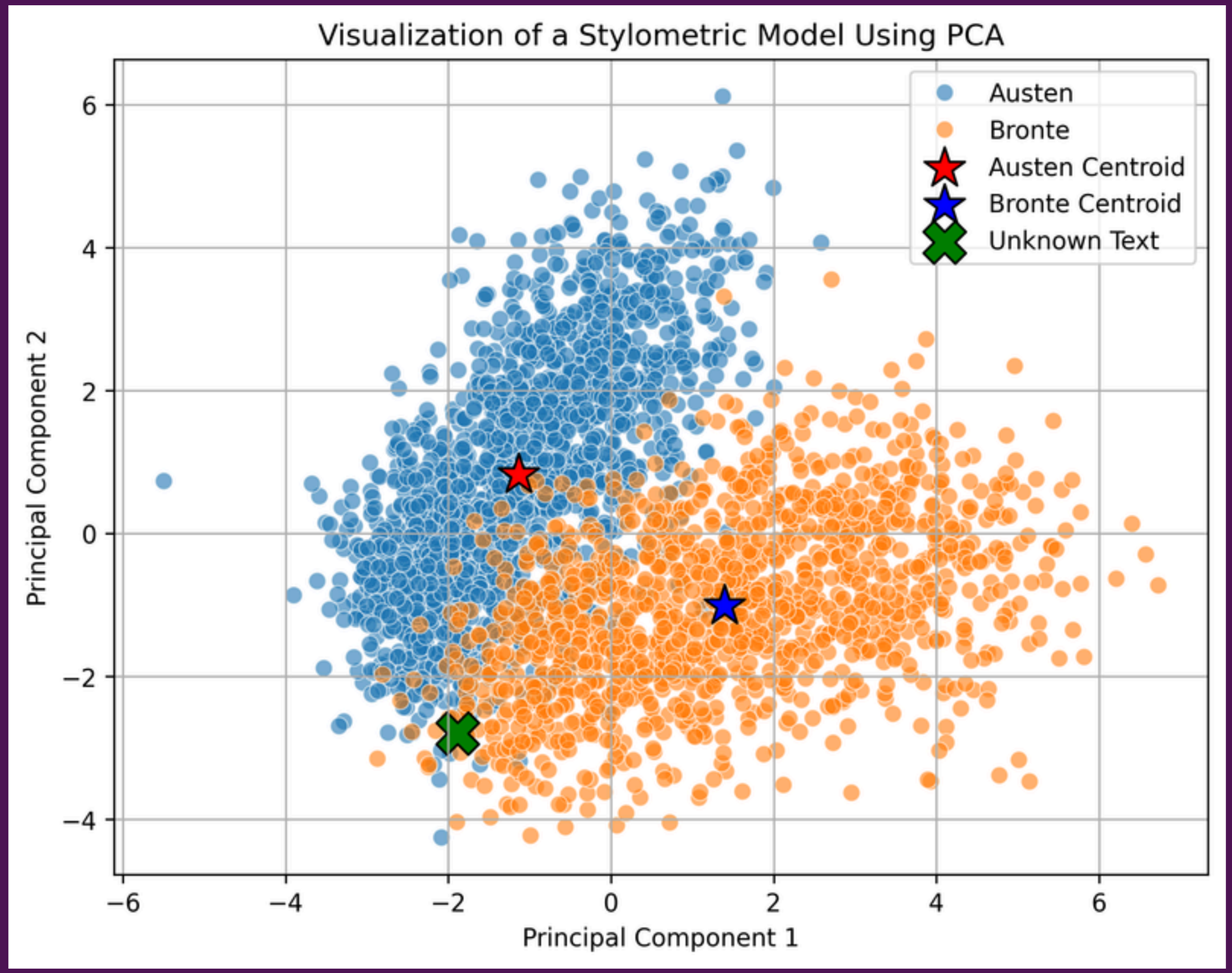
Data Preprocessing

- Regex (re)
- Strategy A:
  - Chunking: 1.000 words
  - Content Masking
- Strategy B:
  - Holding the stop-words
  - max 128 token
  - filling with [PAD]

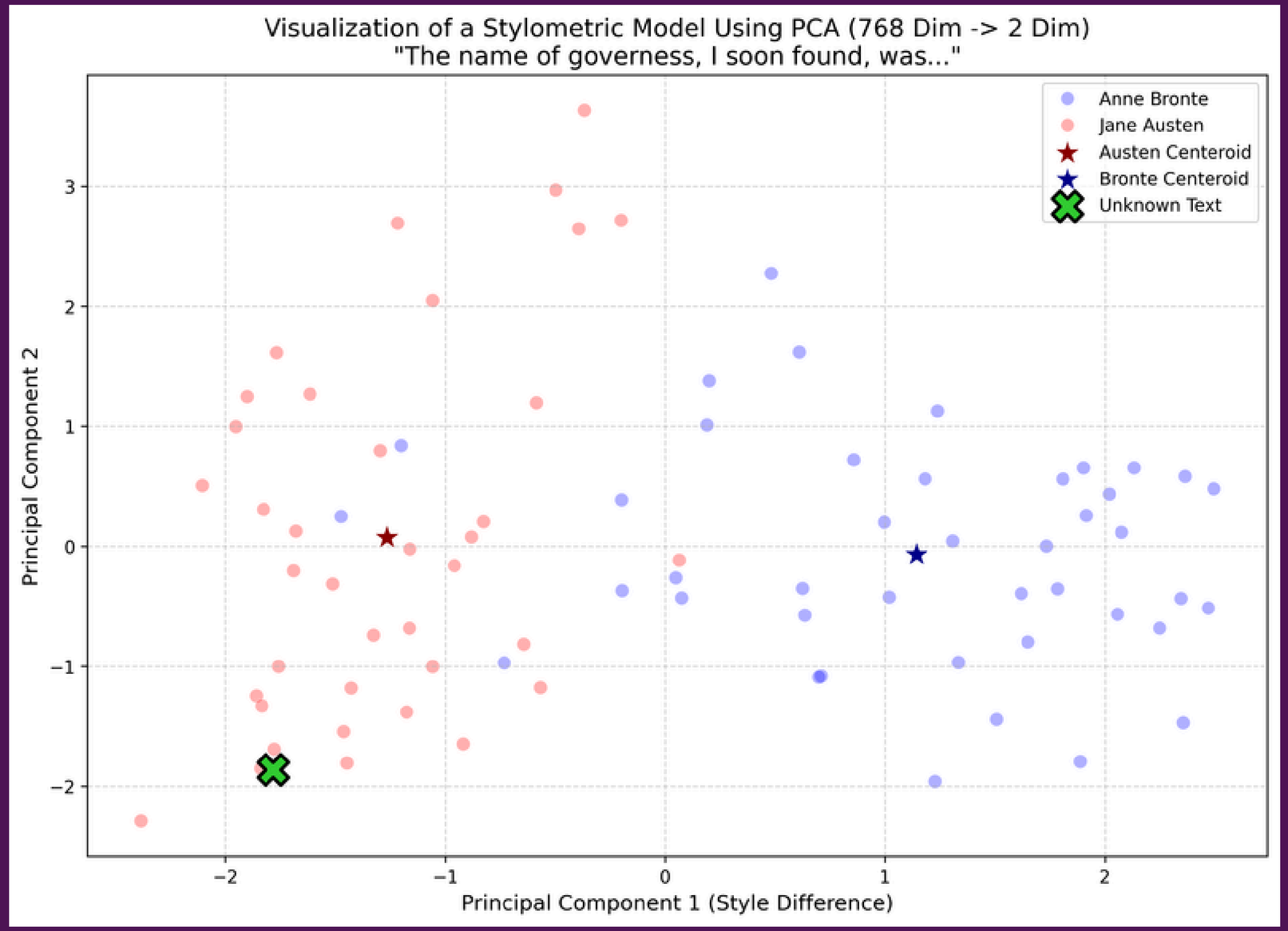
Analysis of Results



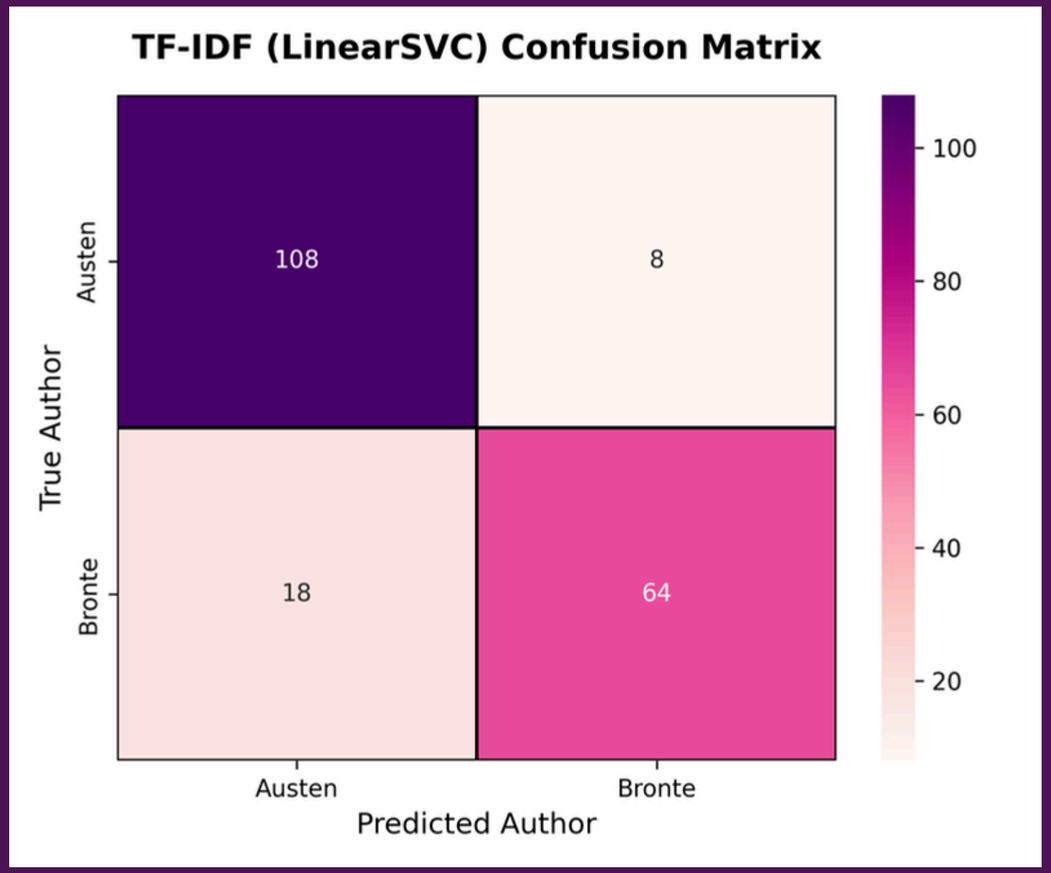
Analysis of Results

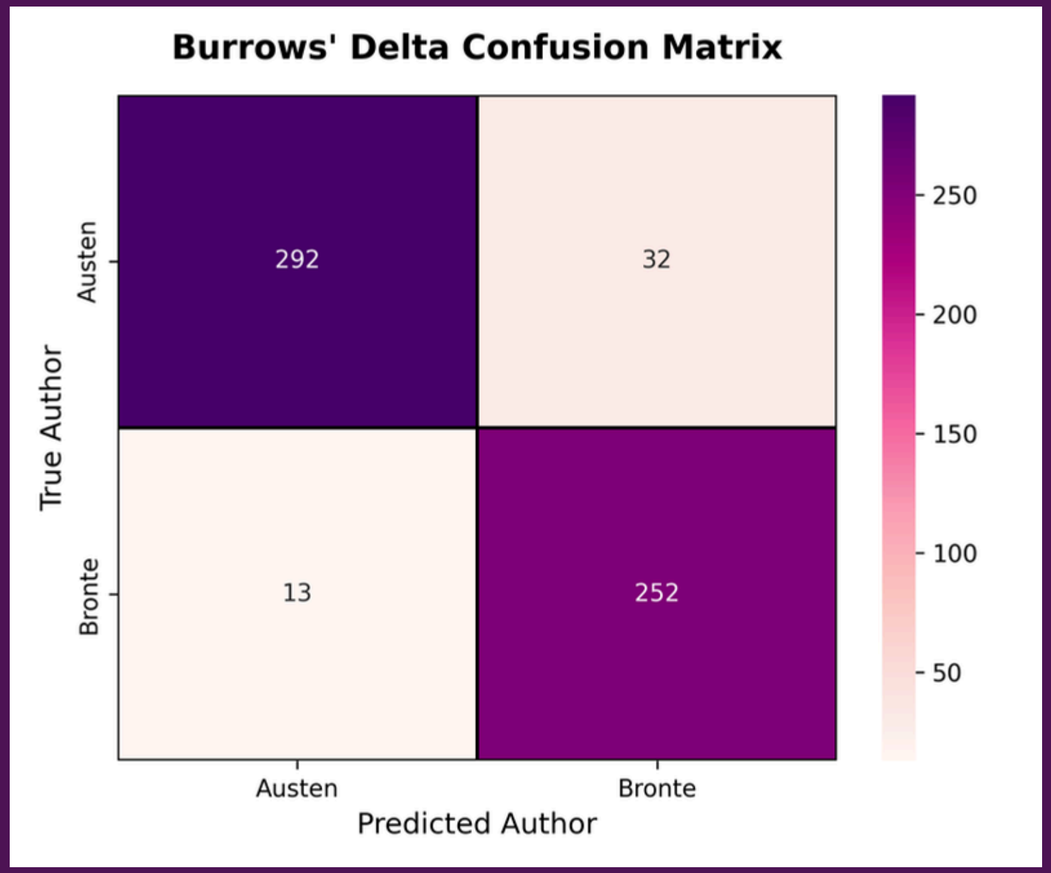


Analysis of Results

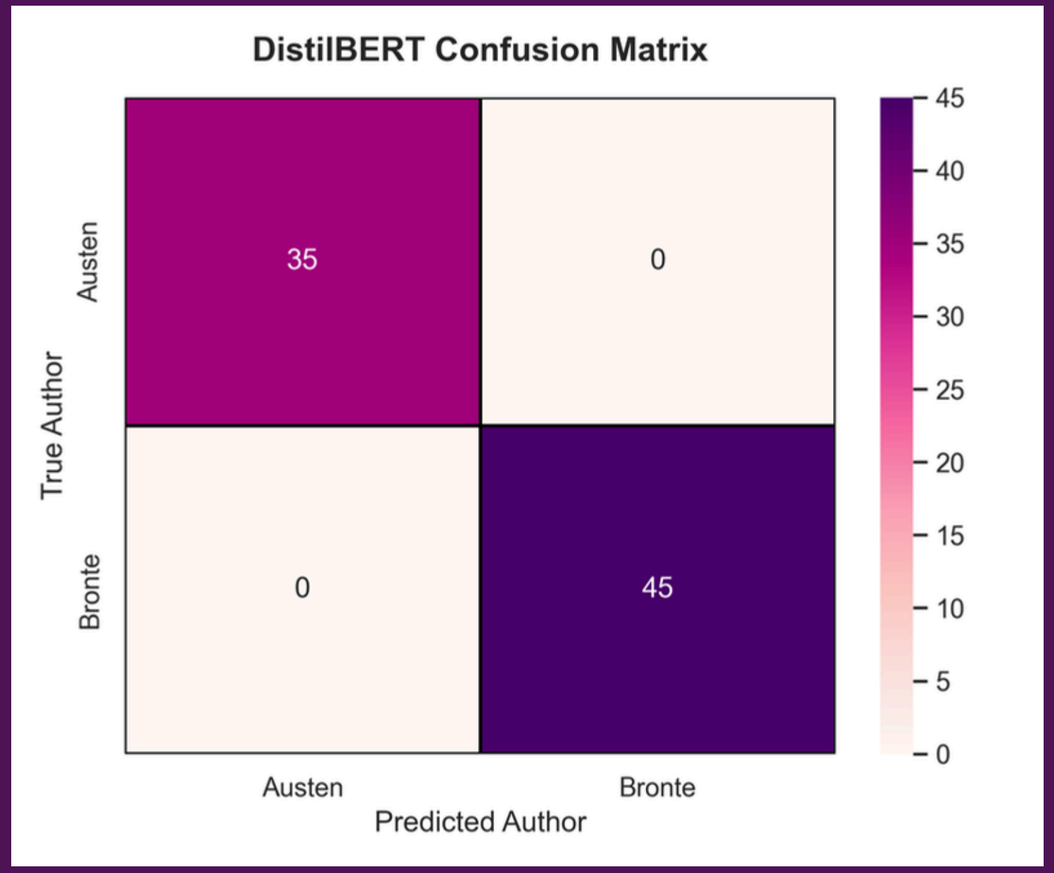


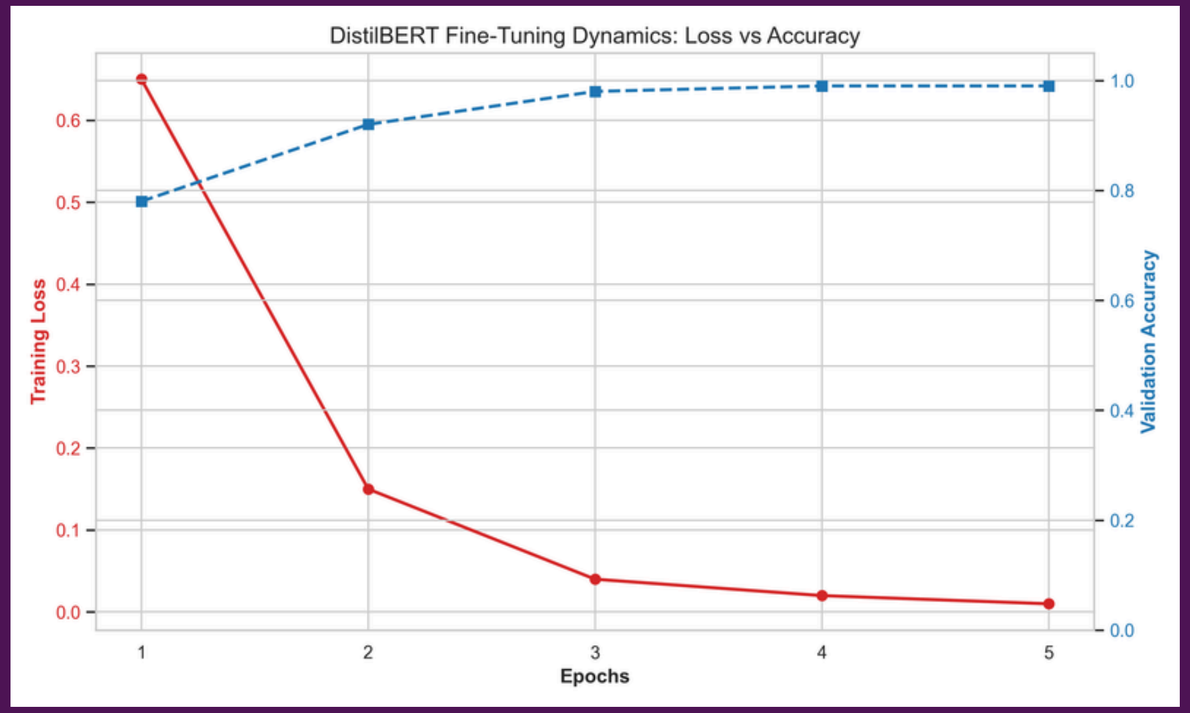
Analysis of Results





Analysis of Results





Conclusion

Model Architecture	Feature Representation	Accuracy	F1-Score	Inference Type
Burrows' Delta	Z-Scores (Top 50 MFW)	0.92	0.92	Distance-Based
LinearSVC (Baseline)	TF-IDF (Function Words)	0.87	0.87	Geometric Hyperplane
DistilBERT (Proposed)	**Contextual Embeddings**	**0.99**	**0.99**	**Semantic Attention**

What’s ahead for agglutinative languages?

Tokenization Problem

- un + believe + able
- yap +??ılan ++ ??dıra ++ ??bildik + ler + imiz

Thank you for your ***attention.***

Any questions?