UNIVERSITY OF LEEDS

جامعة الطائف
TAIF UNIVERSITY
TU

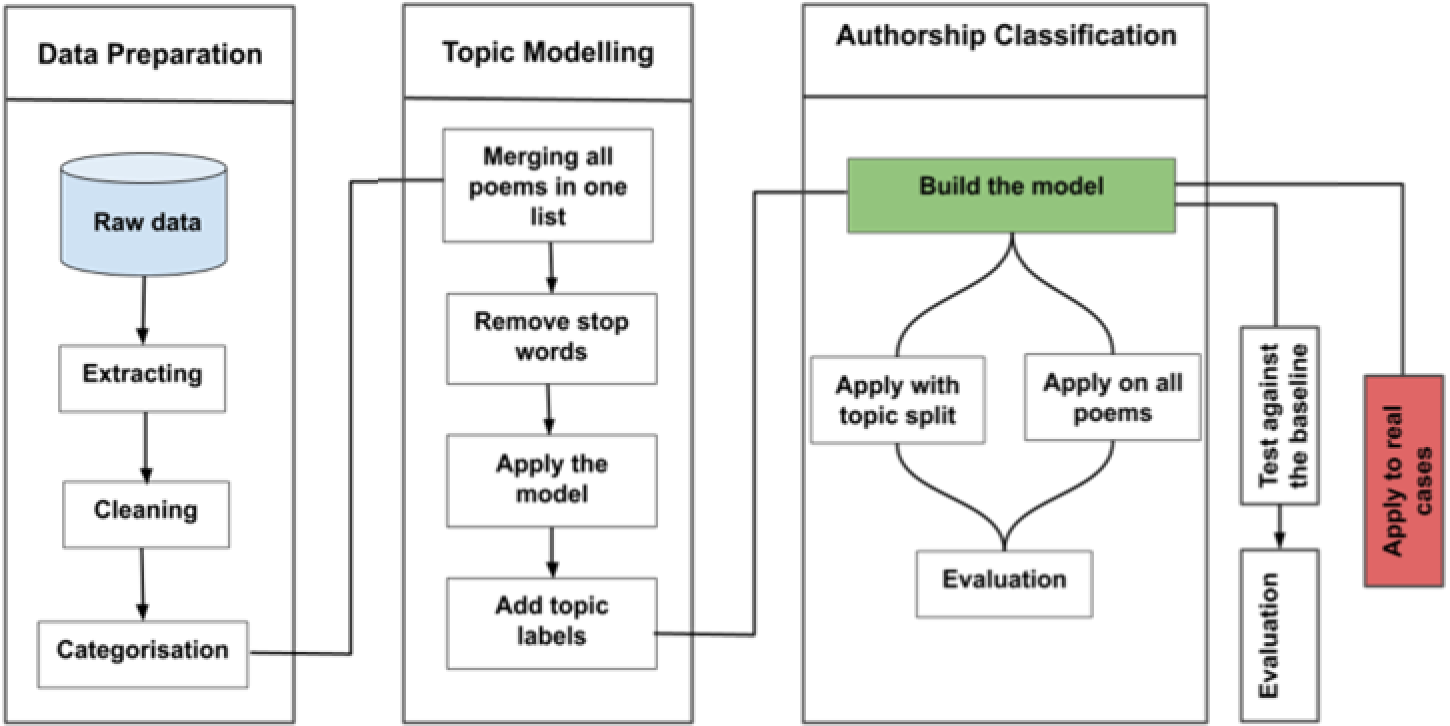
COLING 2025 • Abu Dhabi

The 31st International
Conference on Computational
Linguistics

BERT-based Classical Arabic Poetry Authorship Attribution

Lama Alqurashi, Janet Watson, Jacob Blakesley, Serge Sharoff

Method



Data Preparation

Raw data

Extracting

Cleaning

Categorisation

Topic Modelling

Merging all poems in one list

Remove stop words

Apply the model

Add topic labels

Authorship Classification

Build the model

Apply with topic split

Apply on all poems

Evaluation

Test against the baseline

Evaluation

Apply to real cases

Our Data

Web scraping of **AlDiwan**.etisbew

Category	Total number
Poets	784
Poems	77,850
Words	6,609,495

Topic Modeling

we applied Embedded Topic Modeling (ETM) to label each poem with its topic contributions, further enhancing the dataset’s value. (2020 ,.la te gneiD)

https://github.com/ssharoff/ETM

Authorship Classification

Ensemble model based on **CAMeLBERT** ssorca detset dna depoleved saw :snoisnemid eerht

- Topic
- Number of poets
- Number of training examples

Authorship Classification

After parameter optimization, the model achieved F1 scores ranging from 0.97 to :ta .1.0

- Topic: **No significant effect**
- Number of poets: **Binary**
- Number of training examples: **60**

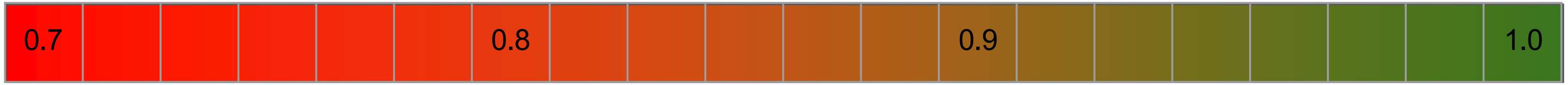
Authorship Classification

The analysis of **average F1** scores for the tested models shows that the ensemble model achieved a score of **0.98**

Significantly outperforming the single model, which obtained a score of .**0.75**

Authorship Classification

We investigated the probability distributions of correctly classified versus misclassified.



Probabilities with **incorrect**noitacifissalc

Probabilities with **correct** classification

Application to Pre-Islamic Mis-attribution Cases

Poem :(syaQ-la 'urmI)1	showed a low confidence score of 0.71
Poem ʿA-lA)2 shā :(	The model's confidence in attribution was 0.86
Poem ʿA-lA)3 shā :(	0.85
Poem -lA nbI dyabU‘)4 Abraṣ :(	high confidence score of ylgnorts ,0.99 noitubirtta sti gnitroppus

أَلا لا أَلَا إِلَّا لآلاءِ لابِثٍ
ولا لَا أَلَا إلا لِآلاءِ مَن رَحَل

Thank you ALL!