Anomaly Detection on Scientific Publications

David-Gabriel ION

Thesis advisor:
Prof. dr. ing. Ciprian DOBRE

THE PROBLEM

Author Identification
for the CRESCDI Platform

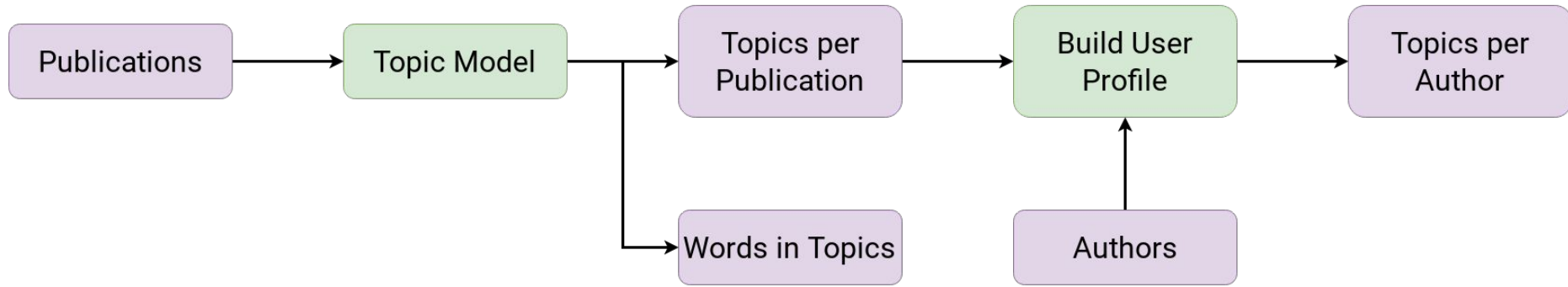# ASSUMPTION

Authors write publications
in specific fields (topics)

IDEA

Using Topic Models
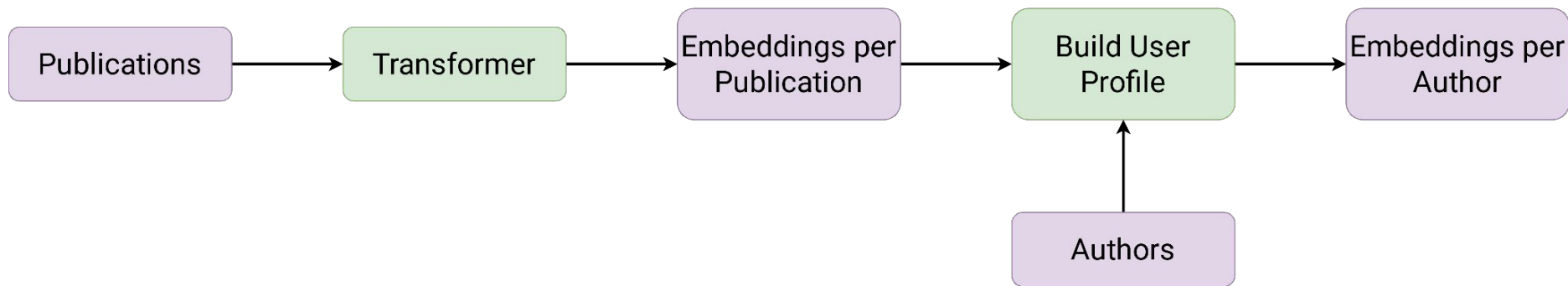
Training

WHAT ABOUT TRANSFORMERS?

Transformers achieve state of the art performance in several NLP tasks
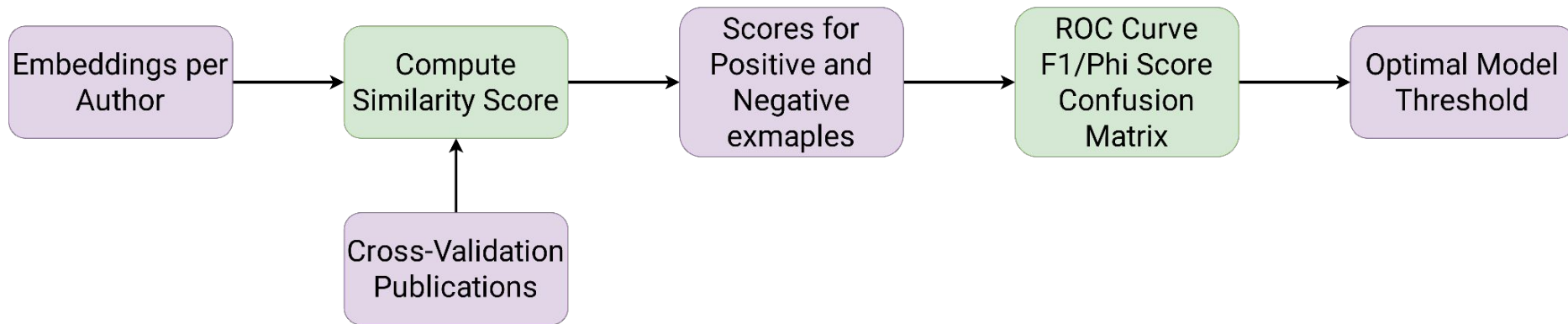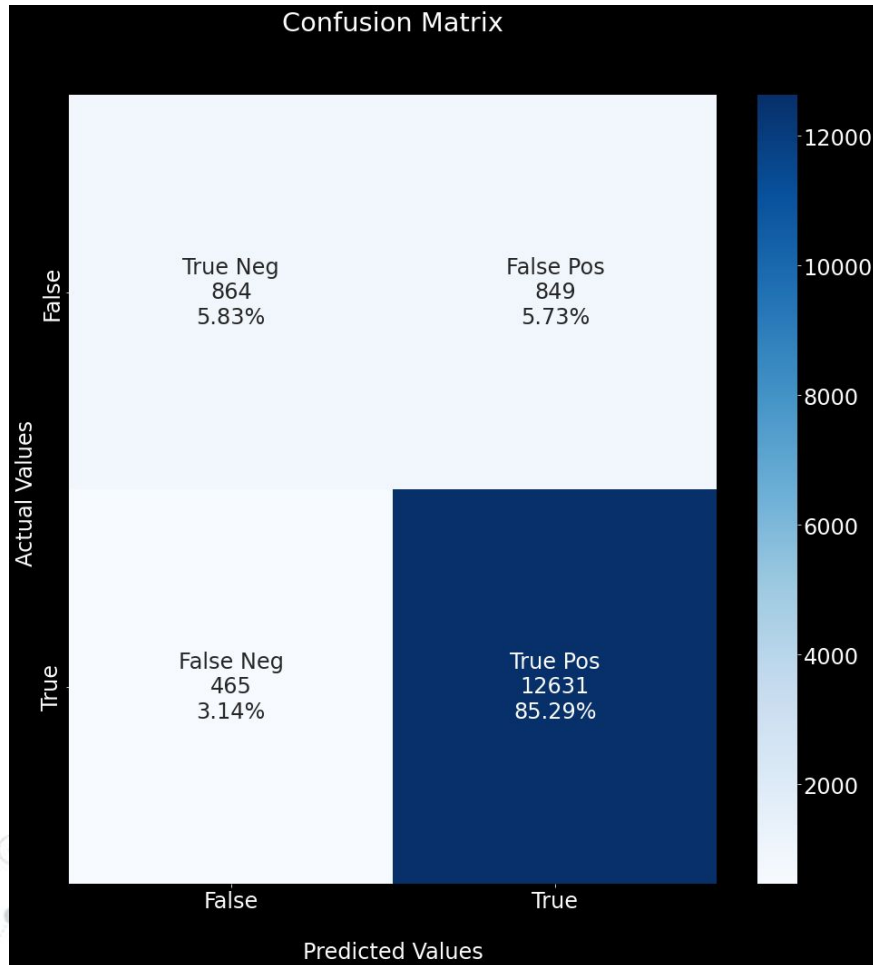
ASSUMPTION

Transformers output similar embeddings
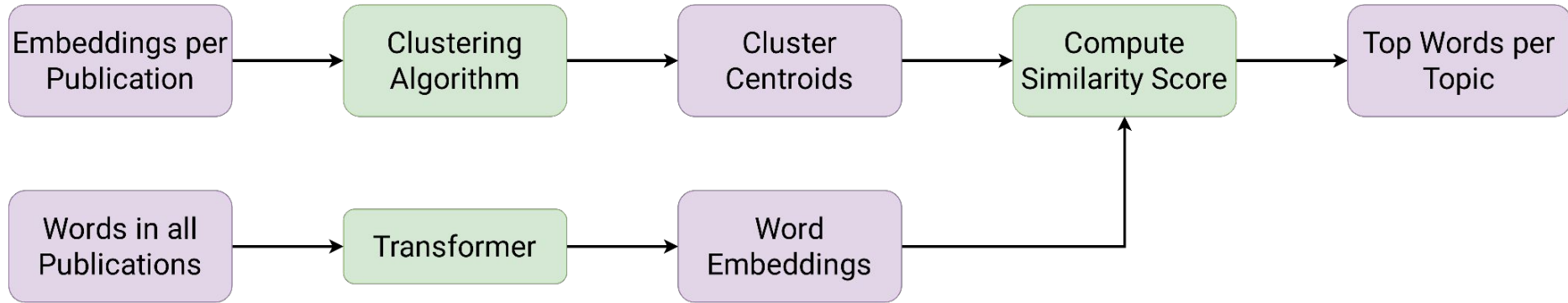for similar documents

Training

Tuning

91% Accuracy
0.95 F1 Score
15 min Train Time
40ms Prediction Time

Topic Modelling

AUTHOR KEYWORDS

Ciprian Dobre
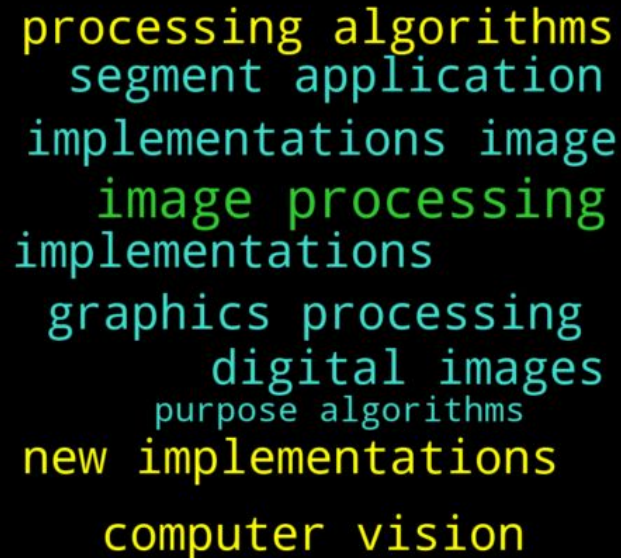
Traian Rebedea

AUTHOR KEYWORDS

Victor Asavei

Cornel Popescu

fin.

Topic modelling
Transformers
Parameter tuning
Keyword extraction