

Feature Engineering

Data Extraction from XML Files

Text Embeddings
X.John et al

Abstract
This paper presents...

1.Introduction
... is important [1]. yyy is...
[1]....[1]....

[1]

ecab3ffd1fe0.xml

① features

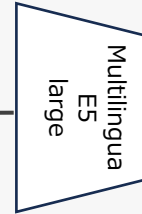
Get ref's abstract



③ citation
count etc.

Generation of Textual Features

Pre text
post text
title
abstract
ref's title
ref's abstract



pair

Vector of a pre text
Vector of post text
Vector of the title
Vector of the abstract
Vector of the ref's title
Vector of the ref's abstract

② Similarity
Cosine similarity and
mahalobius distance
each set



④ Components of UMAP
from text vectors and
node vectors



Vector of
the nodes

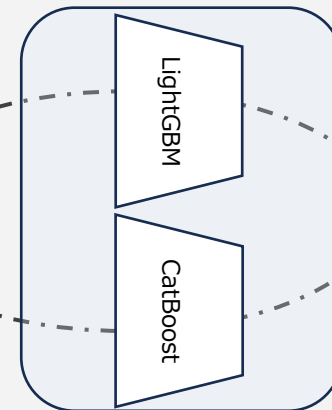
Generation of Network Features

Trainig model and inference

Train Data

① + ② + ③ + ④

Borderline
-SMOTE



Best
ensemble
Model 1

Best
ensemble
Model 2

⋮

Best
ensemble
Model 12

Test Data

① + ② + ③ + ④

y