

# Technical Report: Semantic Scholar Dataset Analysis

Report Generated: 2025-05-05 12:52:24

## Dataset Metadata

- Original records: 1500
- Processed records: 1500
- Processing date: 2025-05-05T12:35:45.698744
- Available columns: title, abstract, authors, journal\_conference\_name, publisher, year, doi, group\_name, processed\_title

## Processing Steps

1. Text cleaning (lowercasing, punctuation removal)
2. Stopword removal
3. Lemmatization
4. Title processing ('processed\_title' column added)

## Dataset Visualizations



No citation statistics available in the dataset