TF-IDF Comparison Summary

# Differences: Manual vs Sklearn TF-IDF

|  |  |  |
| --- | --- | --- |
| Aspect | Manual Implementation | Sklearn TfidfVectorizer |
| IDF Formula | log(N / (1 + df)) + 1 | log((1 + N) / (1 + df)) + 1 |
| TF Calculation | count / total\_terms | Raw count then normalized |
| Normalization | None | L2 norm applied |
| Tokenizer | str.split() | Advanced (removes punctuation, etc.) |
| Stop Words | None | Optional (defaults to None) |