

# IS 688 – Milestone 3 Report

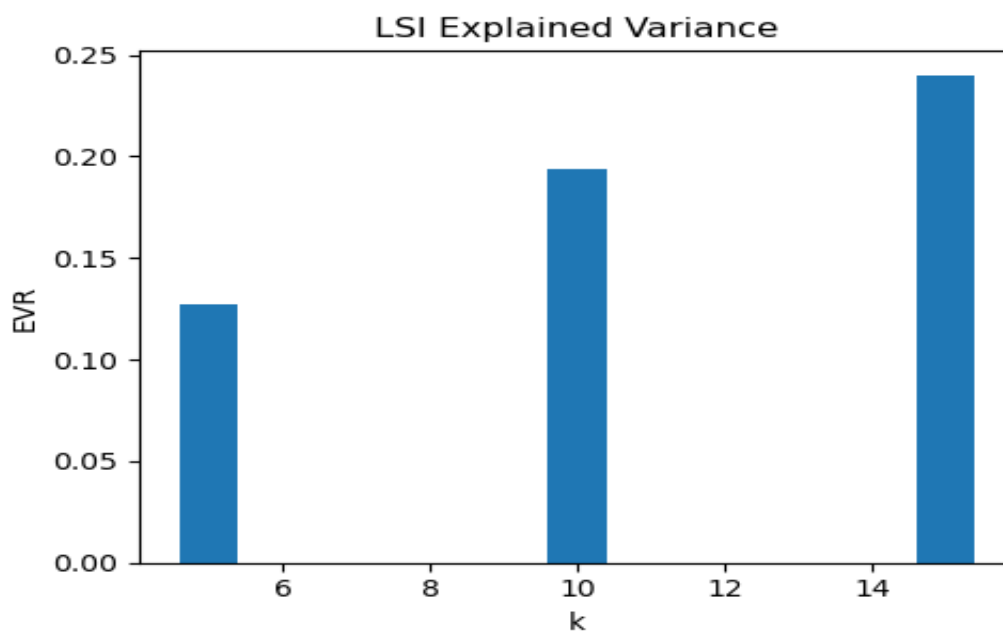
## Topic Modeling with LSI and LDA

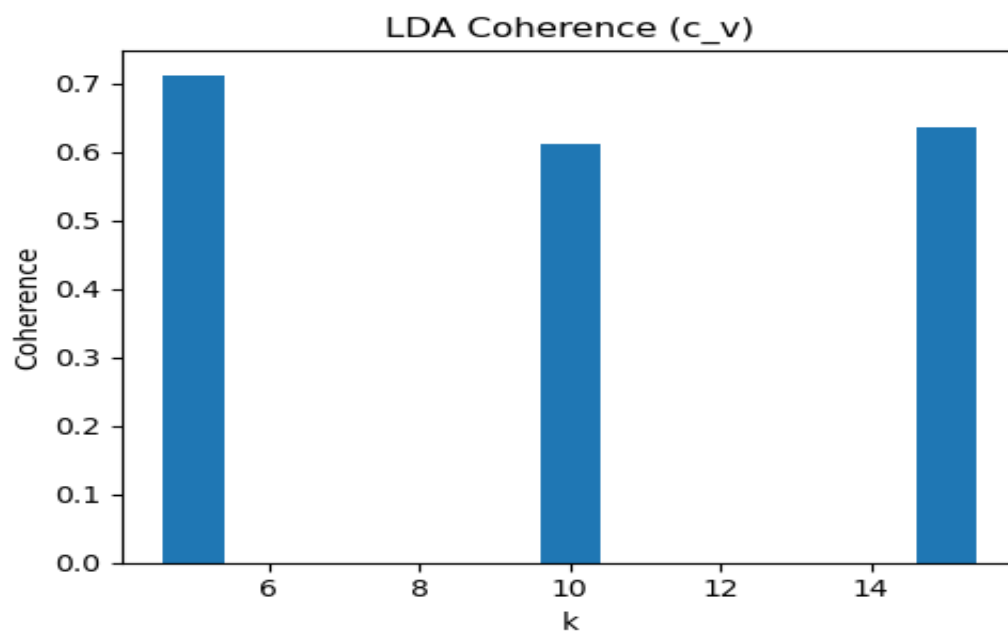
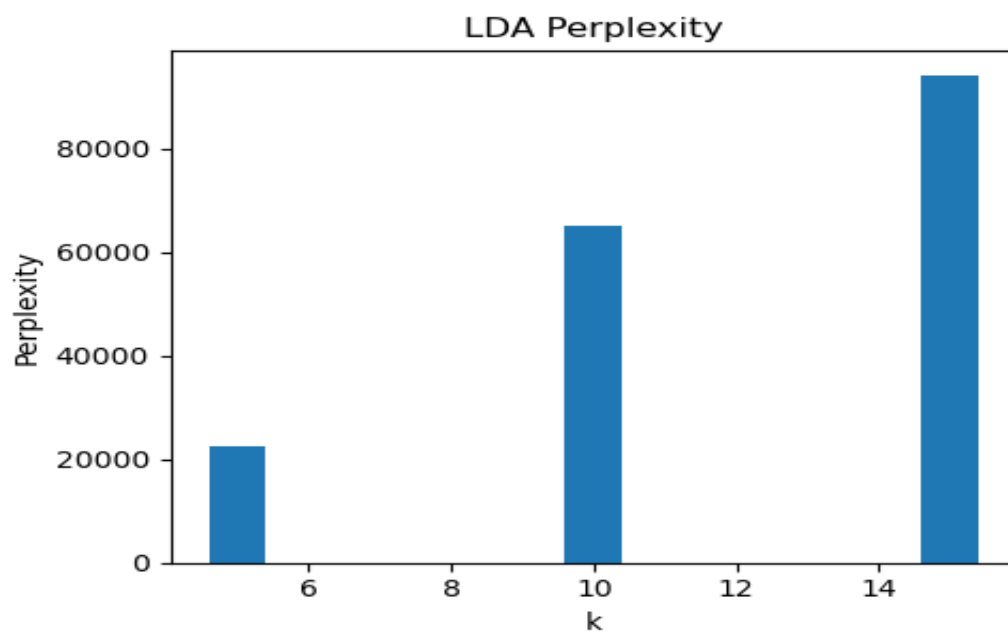
This report summarizes topic modeling experiments using LSI and LDA on the cleaned corpus. LSI was evaluated via explained variance (EVR), while LDA was assessed via perplexity and coherence.

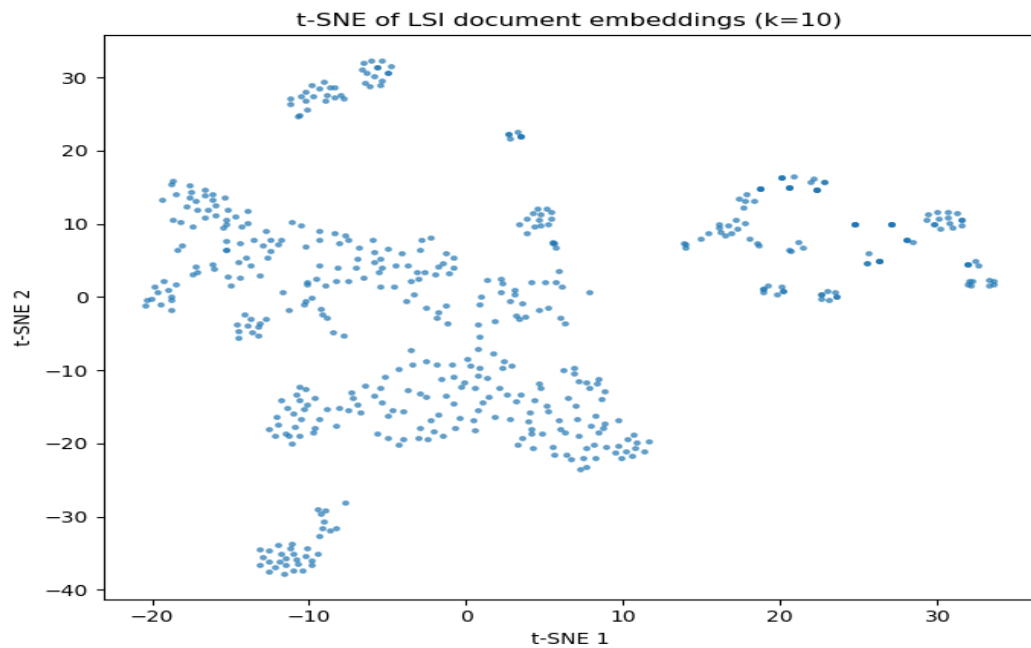
### ***Model Comparison Table***

k	LSI EVR	LDA Perplexity	LDA Coherence (c_v)
5	0.1270	22342.55	0.7148
10	0.1941	64081.19	0.6116
15	0.2400	94064.52	0.6355

LSI with  $k=15$  captured the richest structure ( $EVR=0.24$ ), while LDA with  $k=5$  provided the best coherence (0.71). Increasing  $k$  raised perplexity, indicating overfitting. Both methods reveal interpretable topic clusters aligned with computational and mathematical themes.







Generated automatically via Python (ReportLab) | © 2025 Pierre Moise