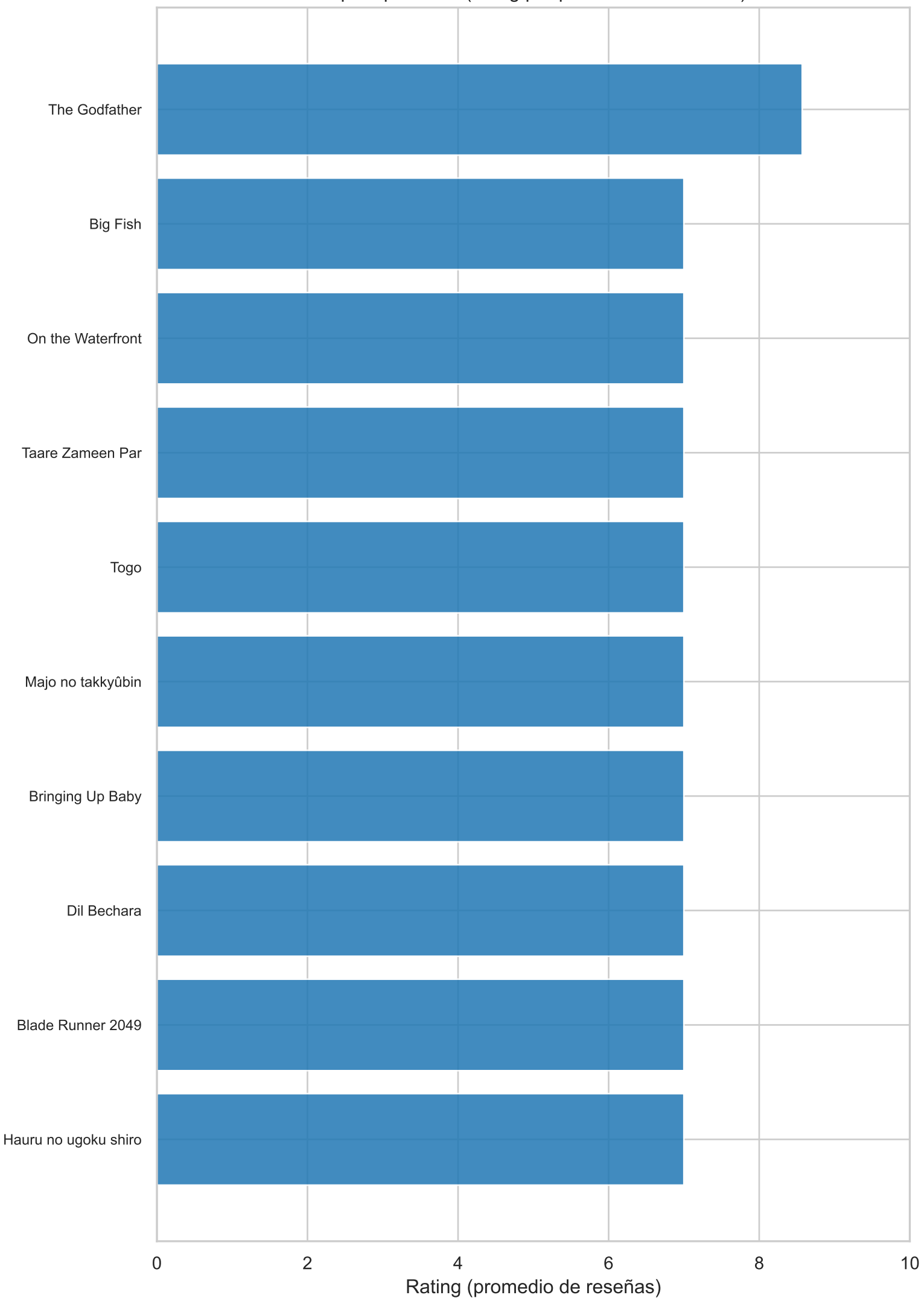


Top 10 películas (rating por promedio de reseñas)



IMDB - Model Results

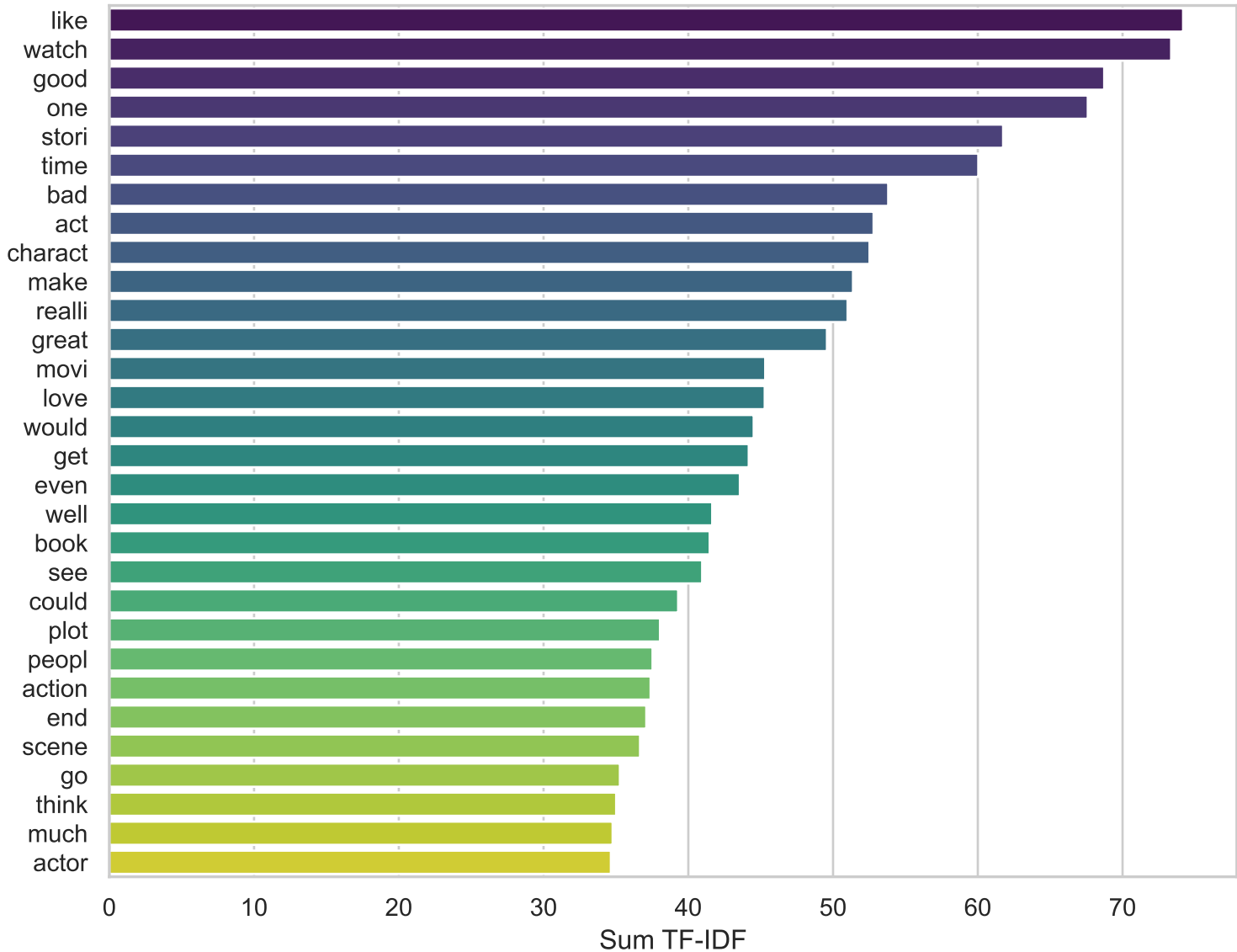
Total reviews in DB: 5458

Reviews used for modeling (first 4000): 3997

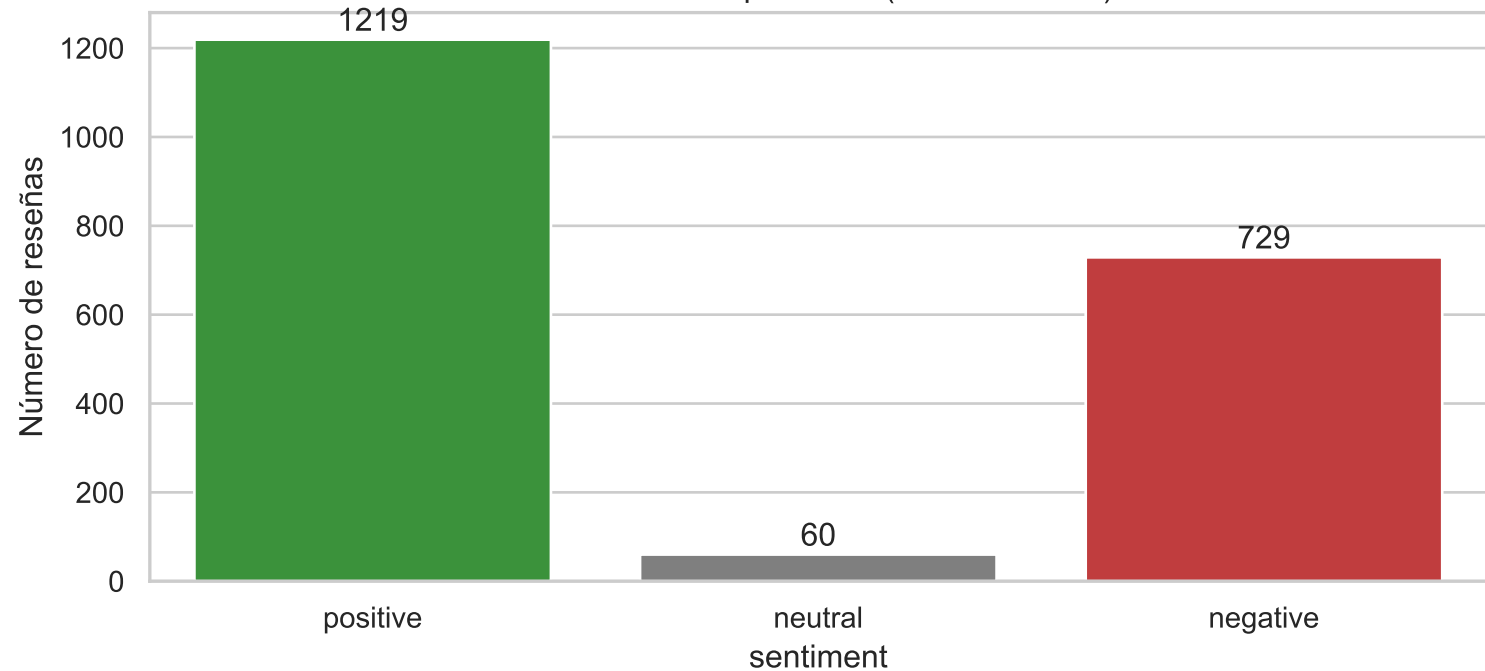
Reviews used for PDF visuals/evaluation (JSON): 2008

Models trained: Naive Bayes, Linear SVM, Decision Tree

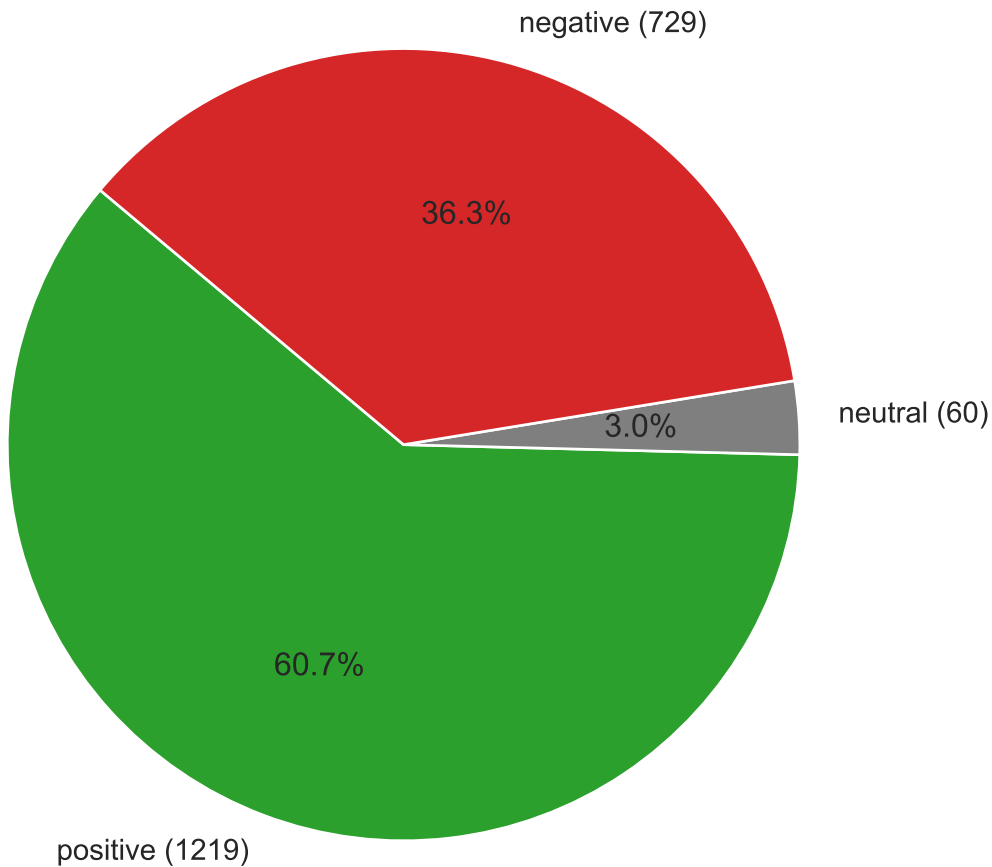
Top TF-IDF terms (reviews from JSON)



Distribución de polaridad (reseñas JSON)

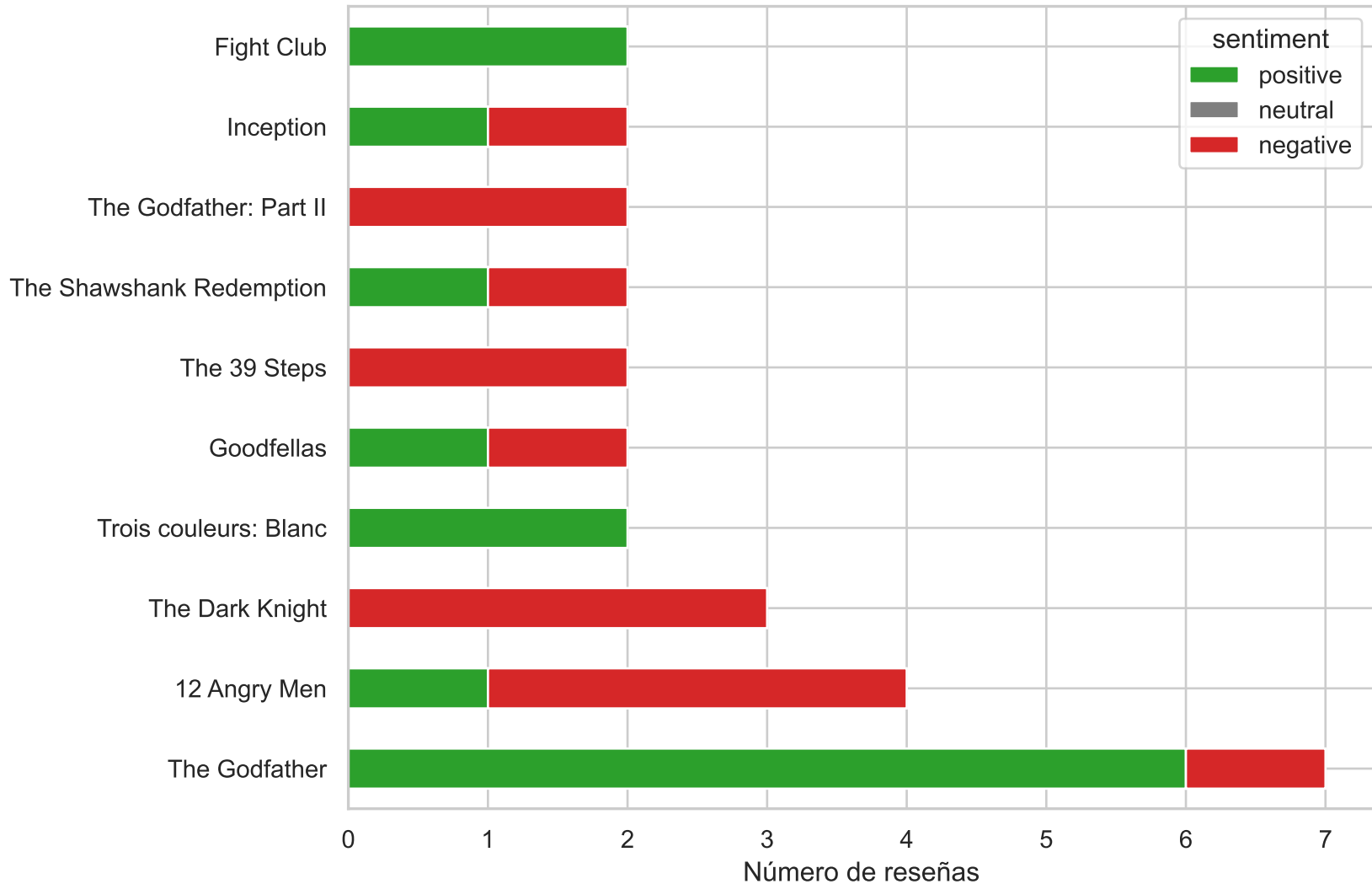


Porcentaje de polaridad (reseñas JSON)

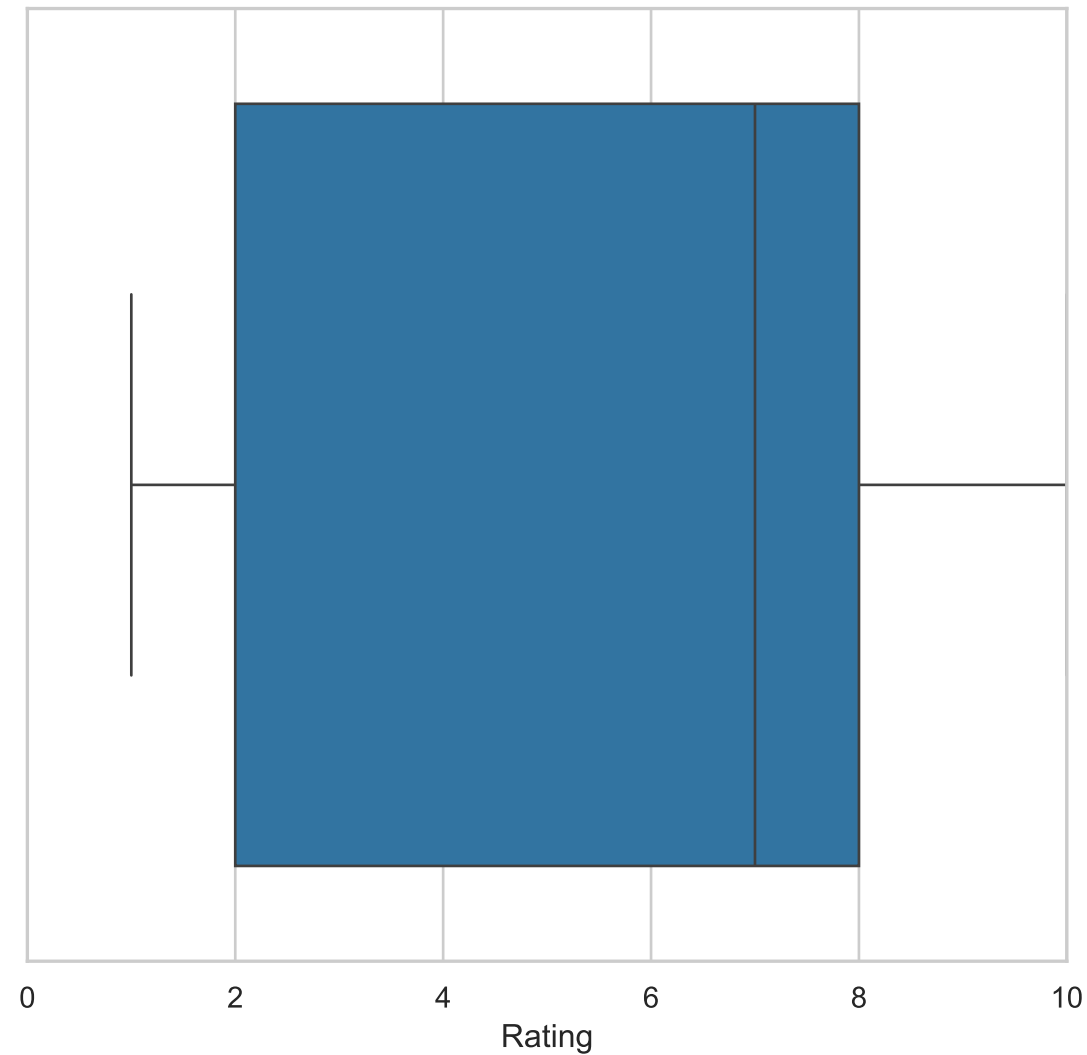


Polaridad por película (top 10 por #reseñas en JSON)

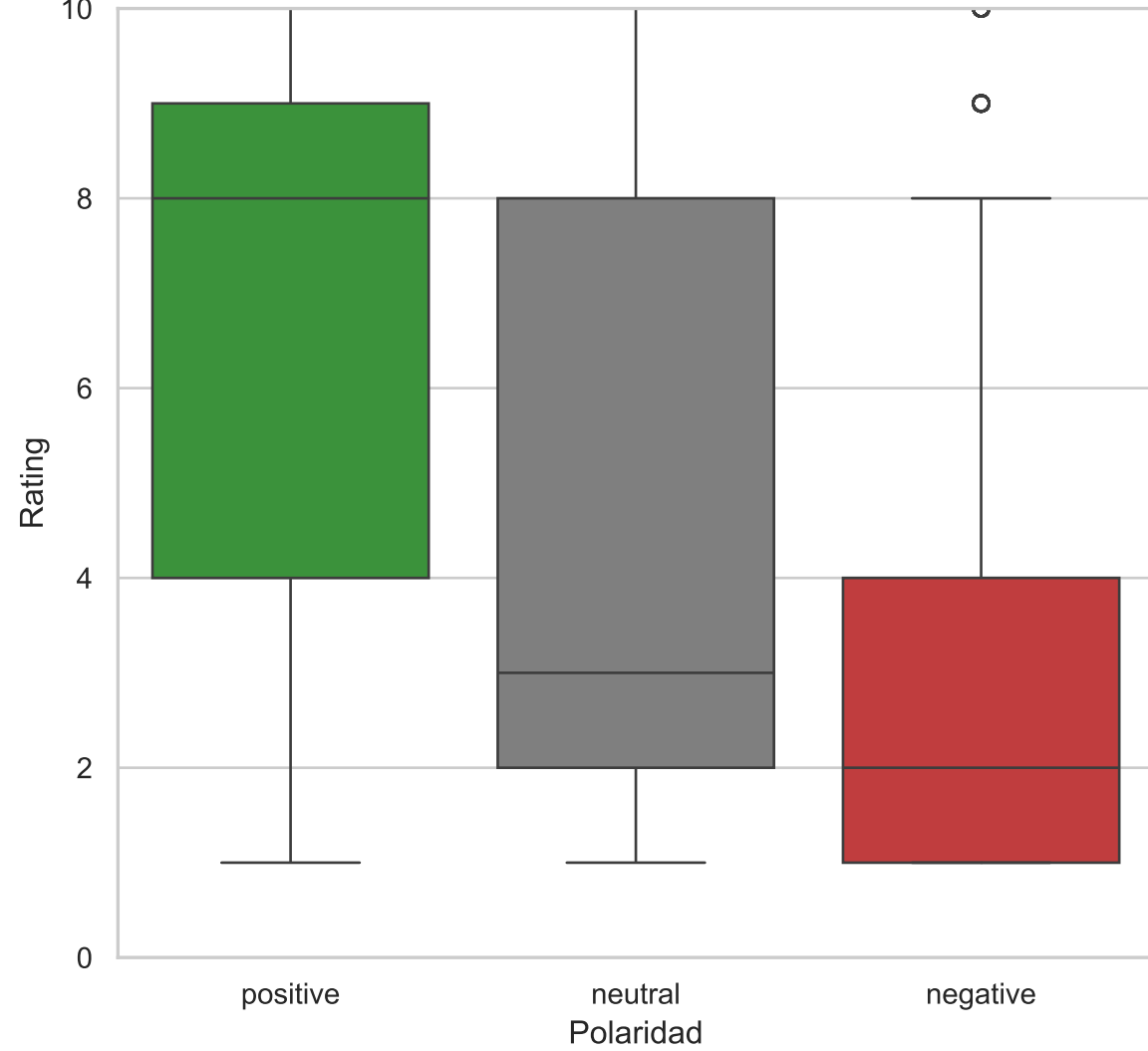
Película



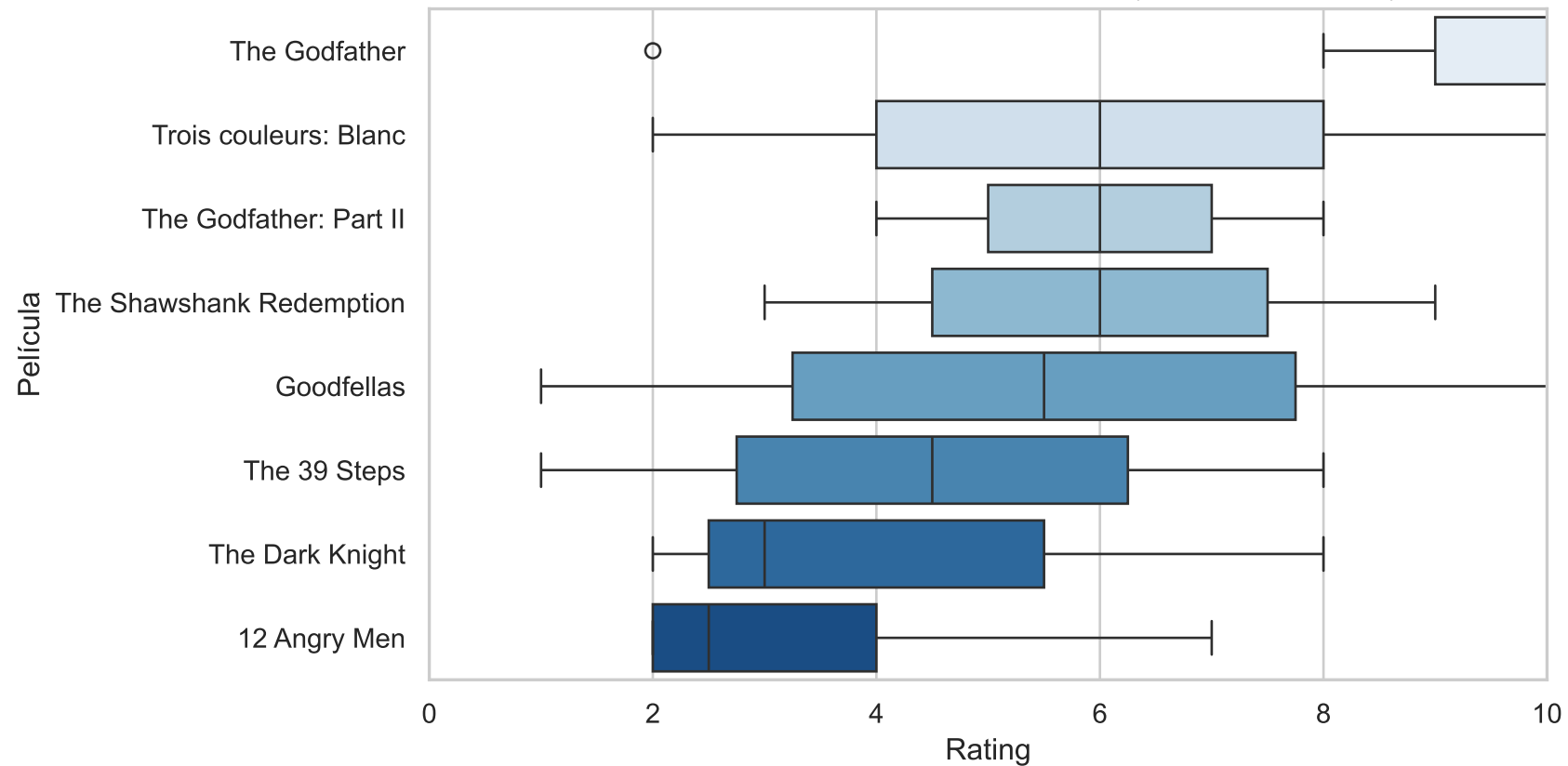
Distribución de calificaciones (boxplot) - global



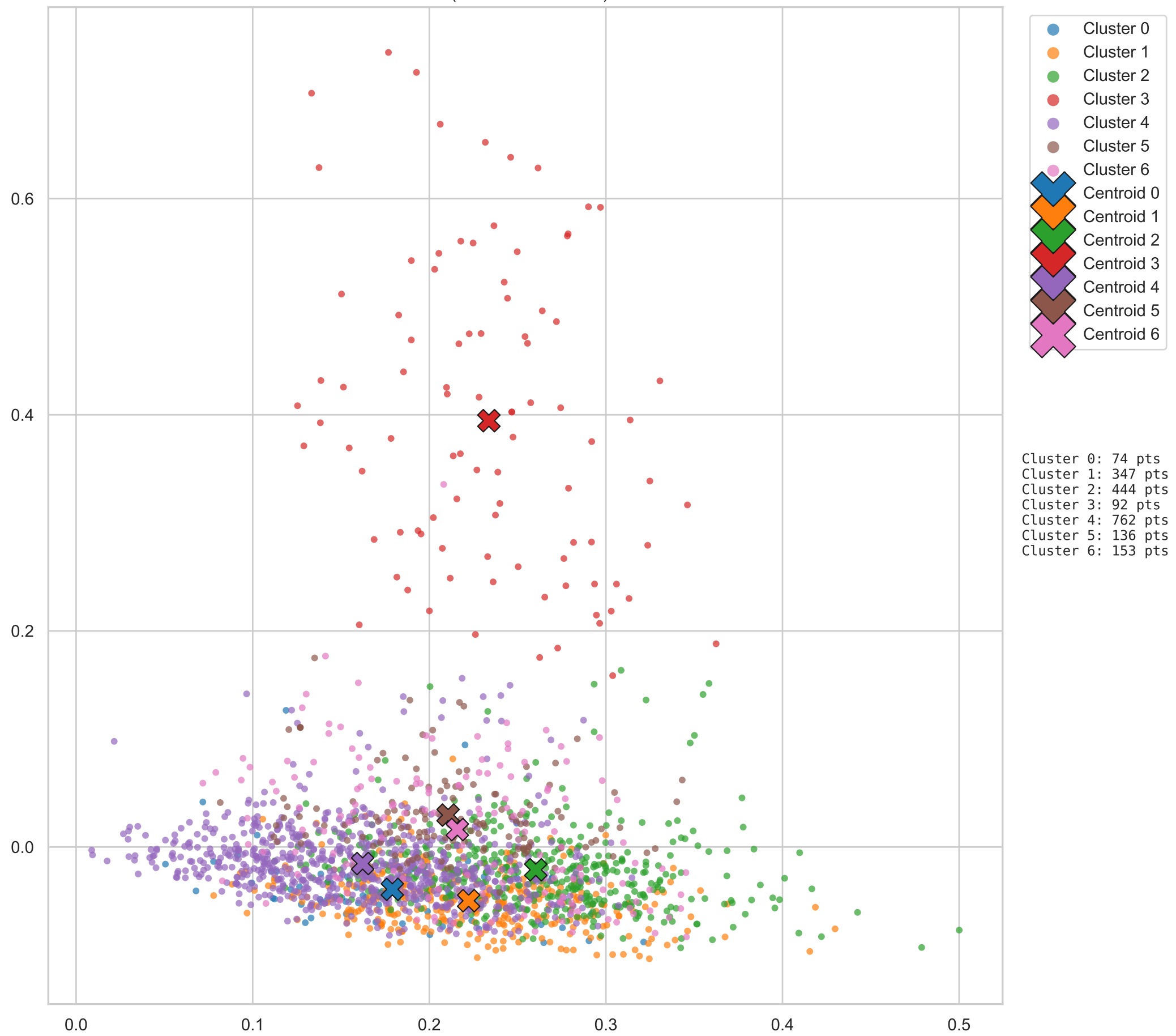
Distribución de calificaciones por polaridad (boxplot)



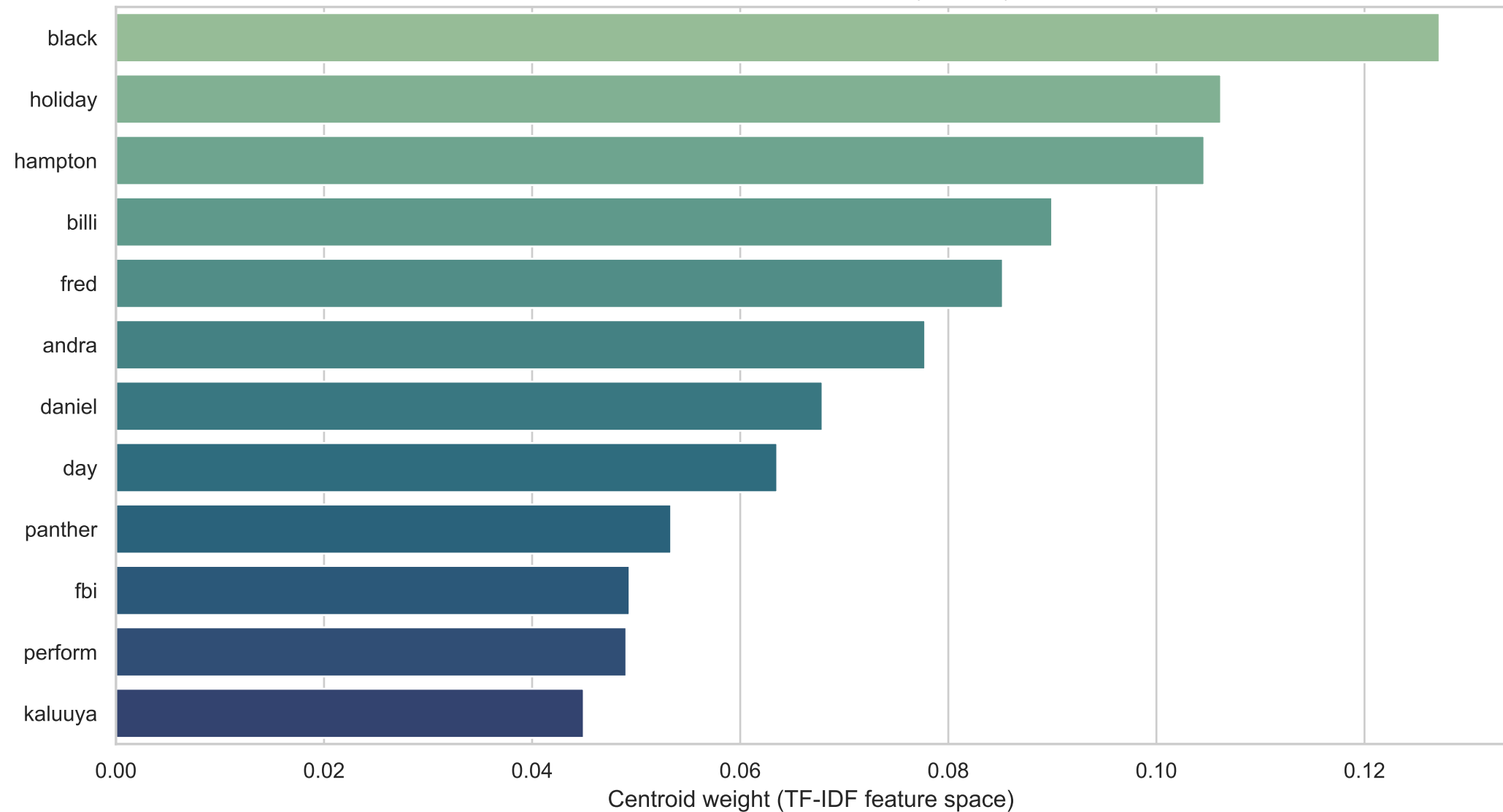
Distribución de calificaciones por película (top 8 por #reseñas)



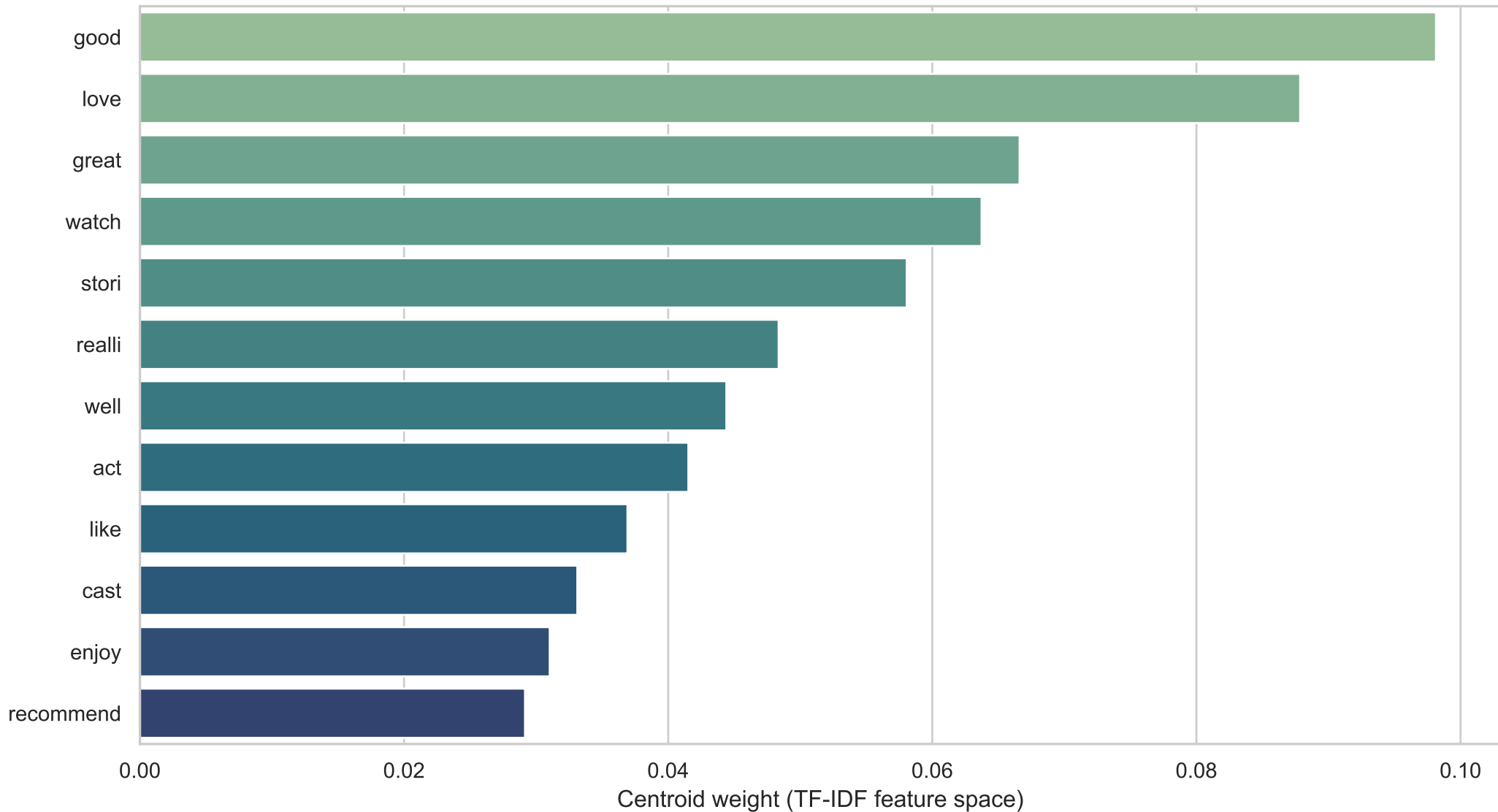
KMeans clusters (TruncatedSVD 2D) - JSON reviews



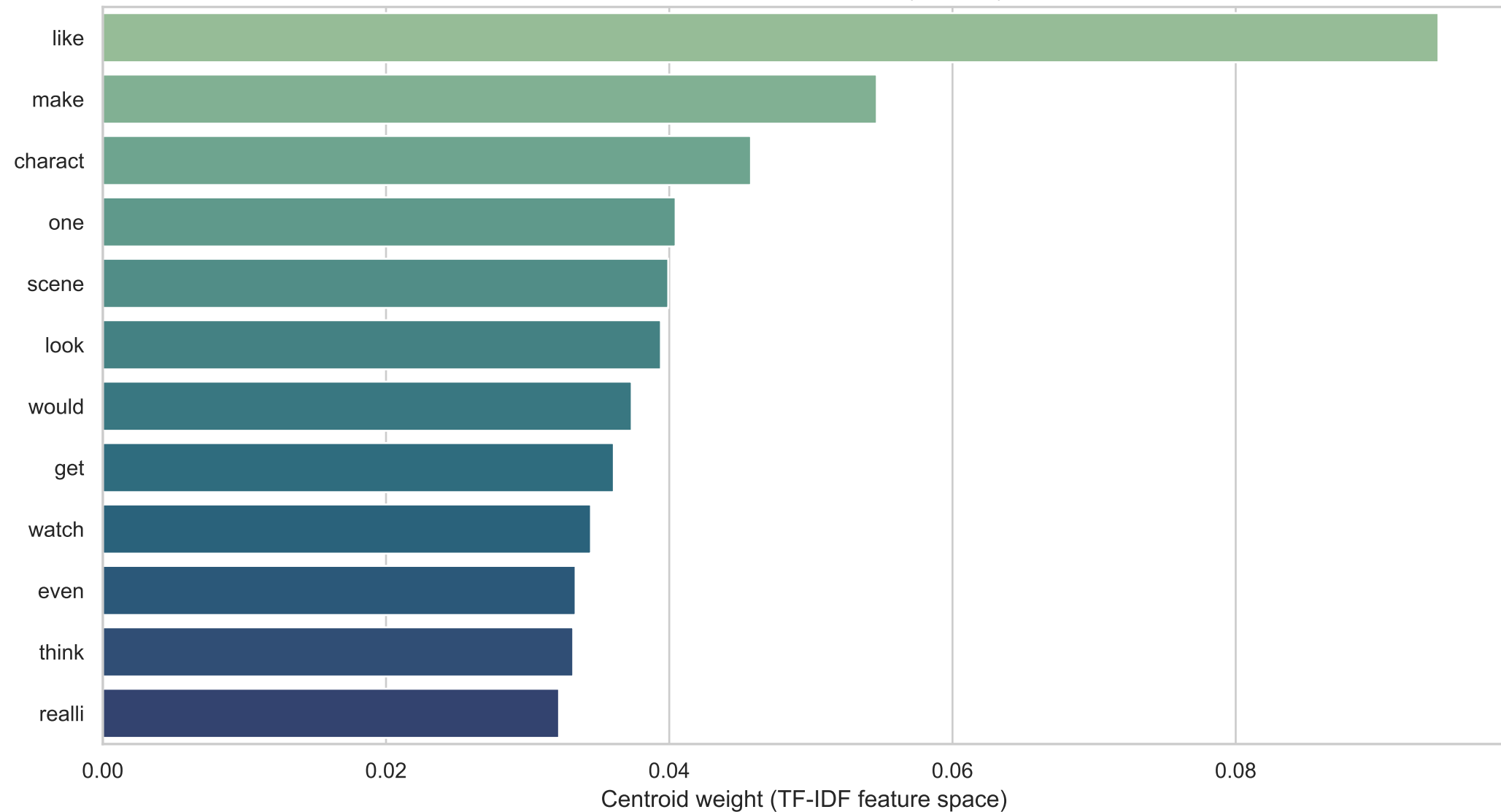
Cluster 0 - top TF-IDF words (centroid)



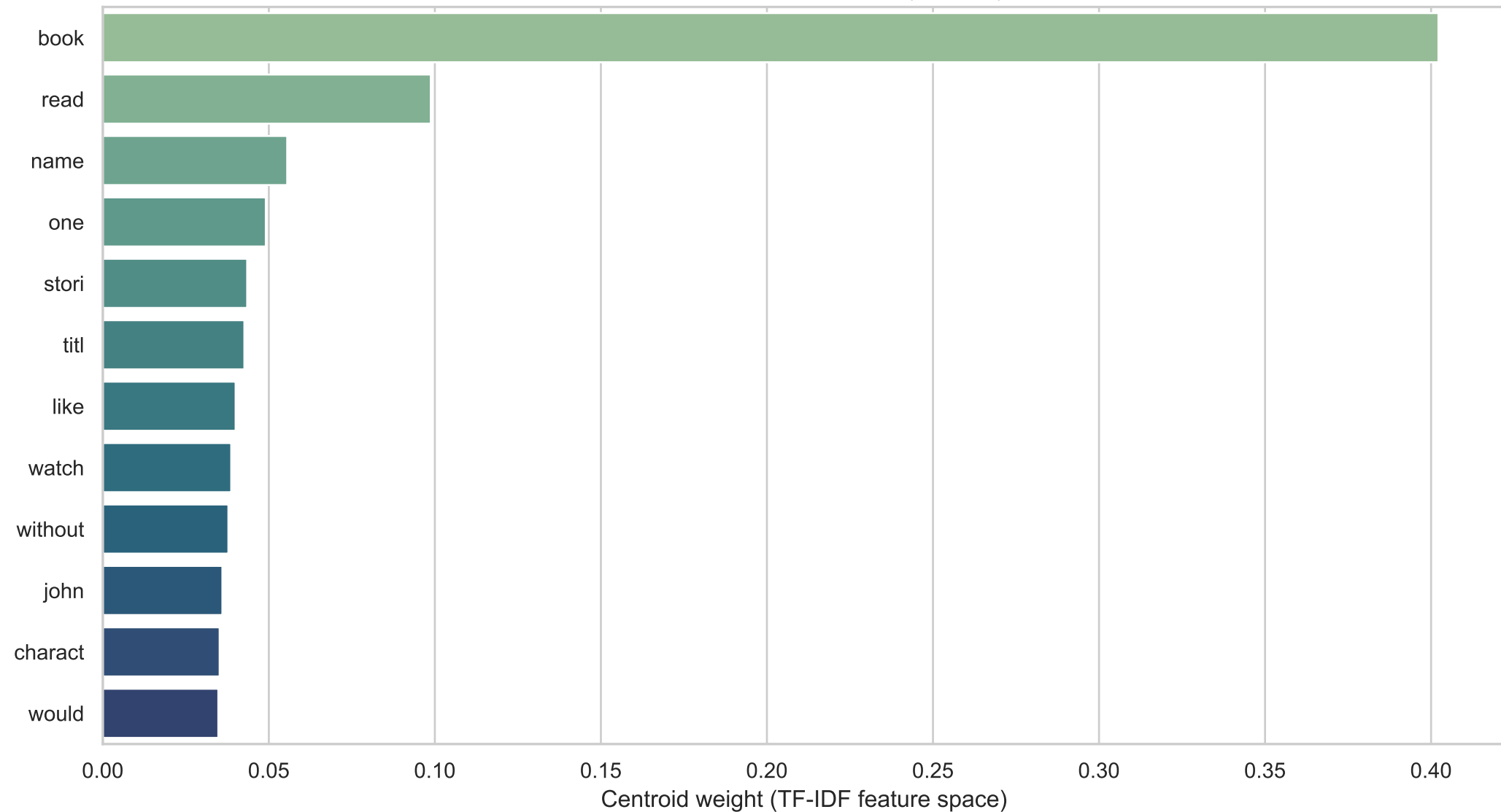
Cluster 1 - top TF-IDF words (centroid)



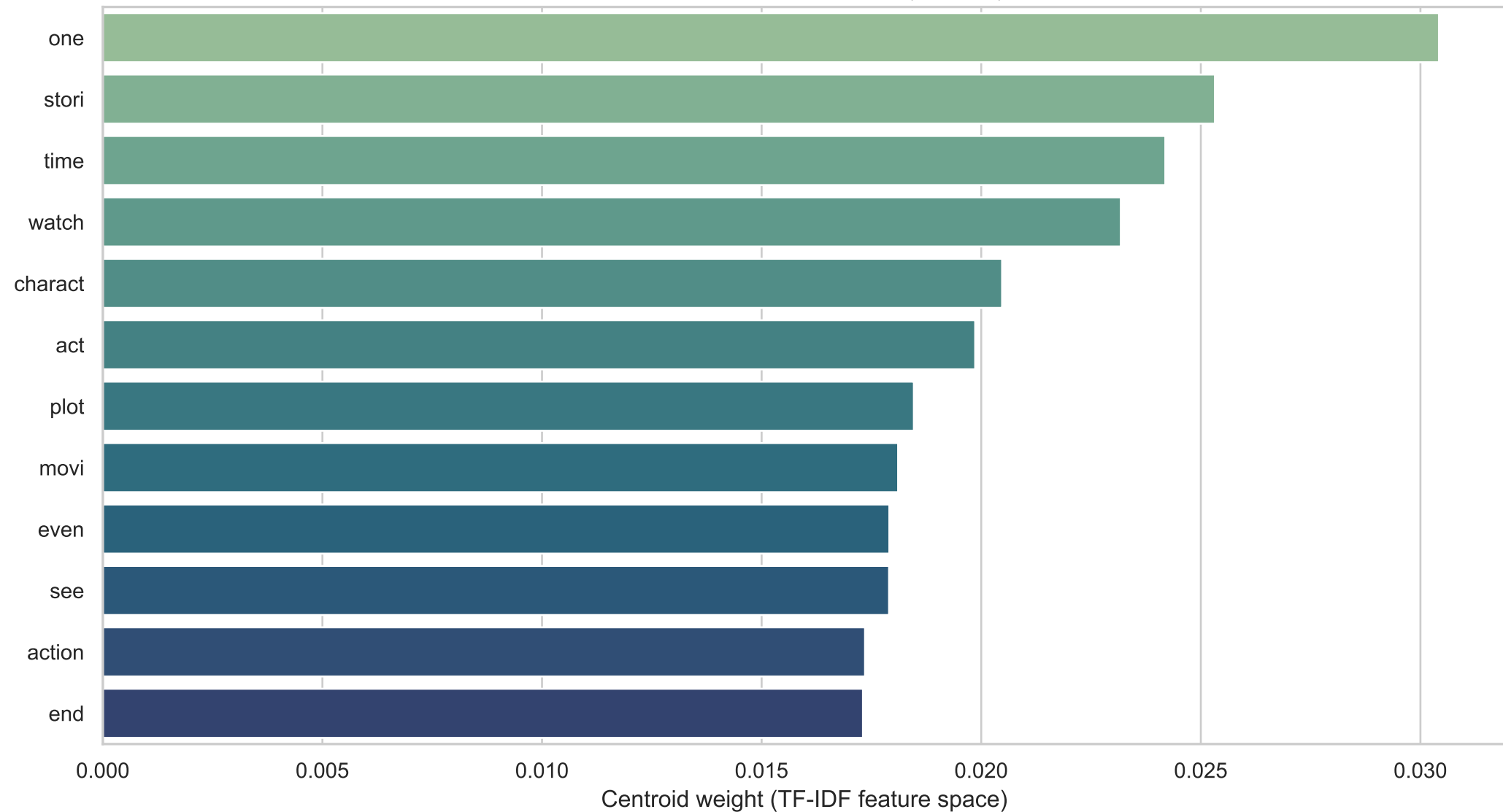
Cluster 2 - top TF-IDF words (centroid)



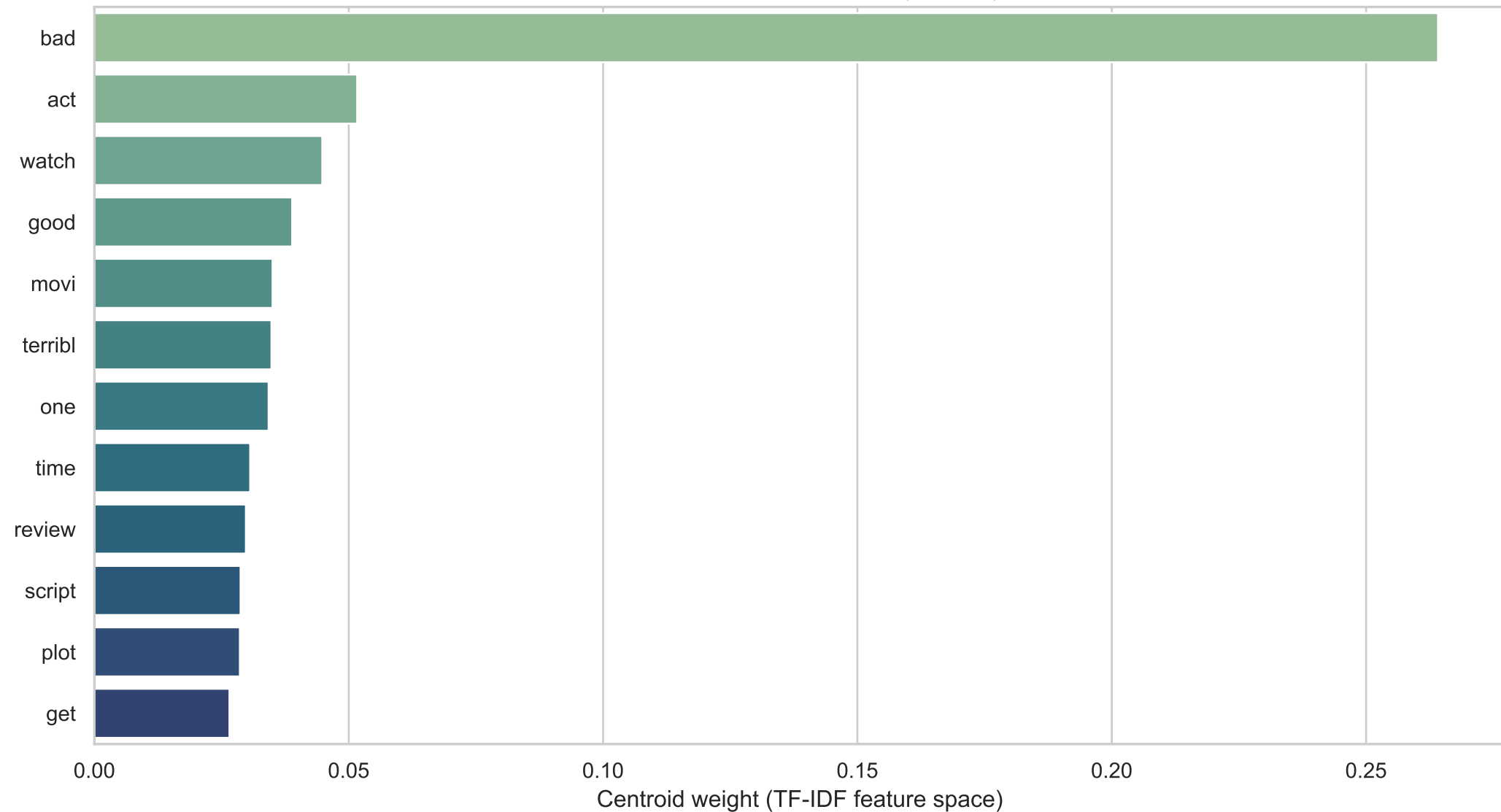
Cluster 3 - top TF-IDF words (centroid)



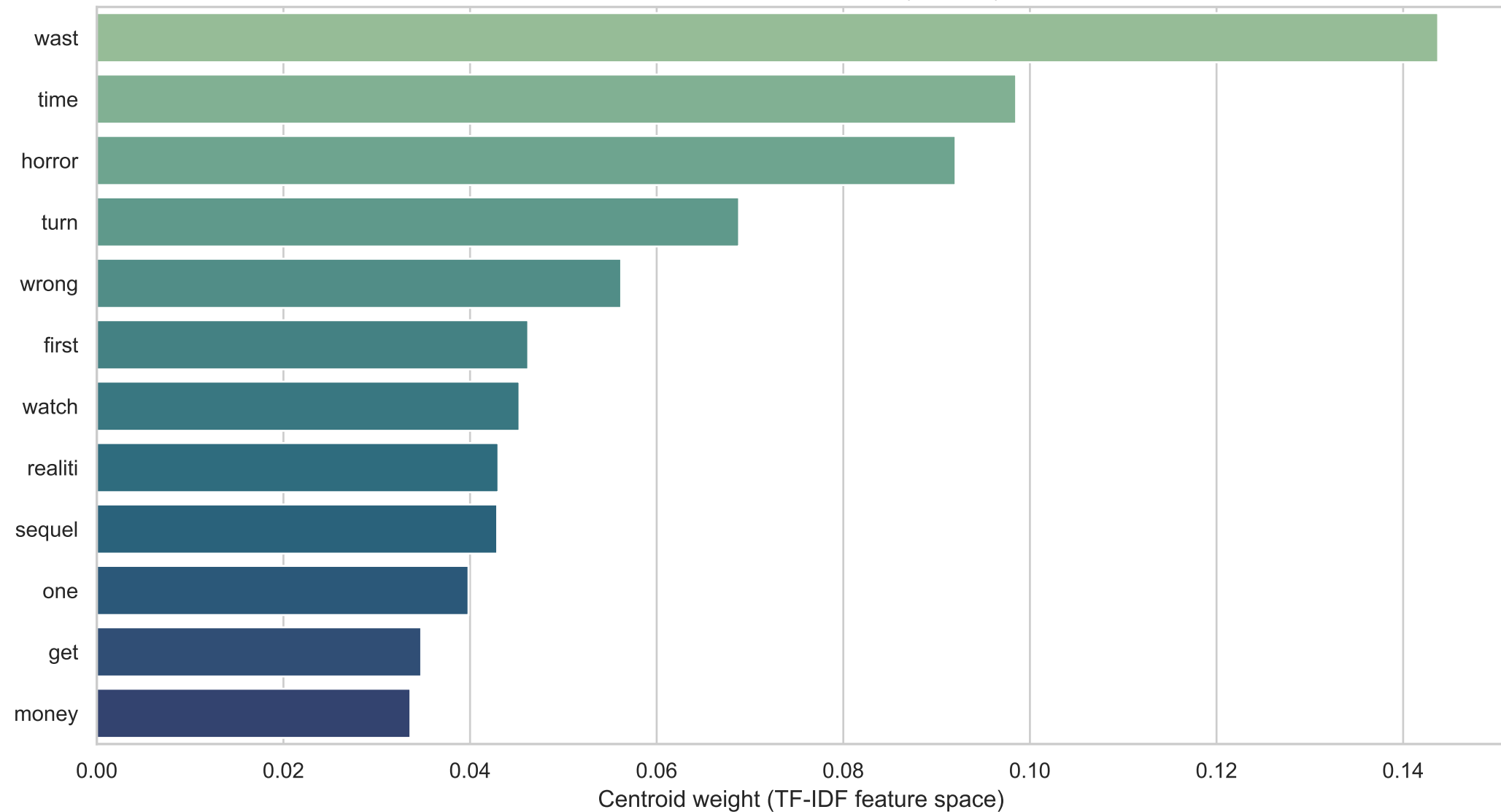
Cluster 4 - top TF-IDF words (centroid)



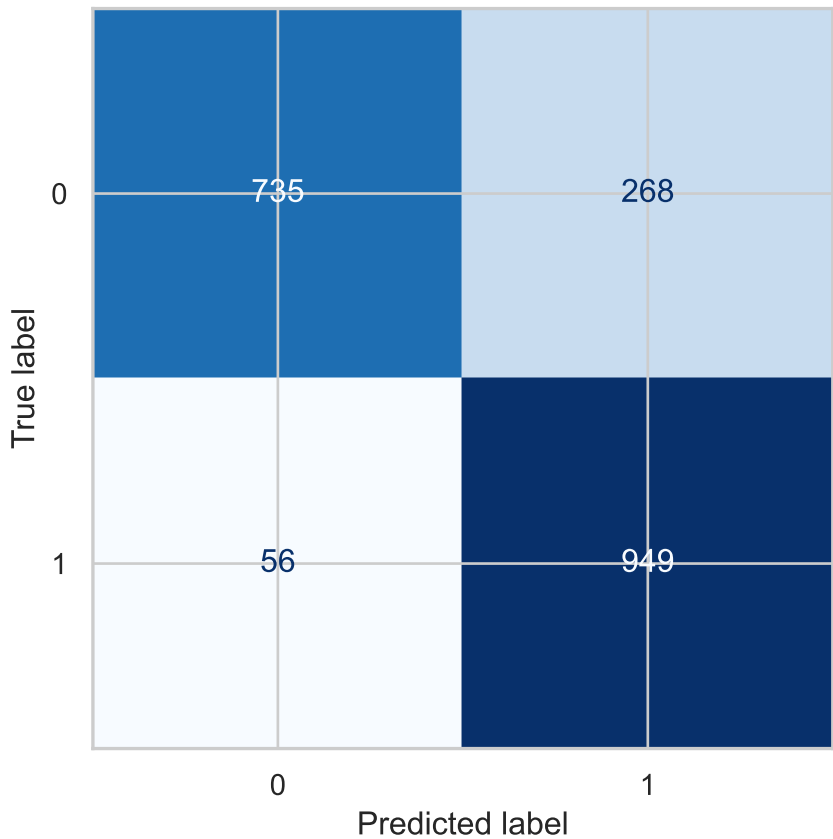
Cluster 5 - top TF-IDF words (centroid)



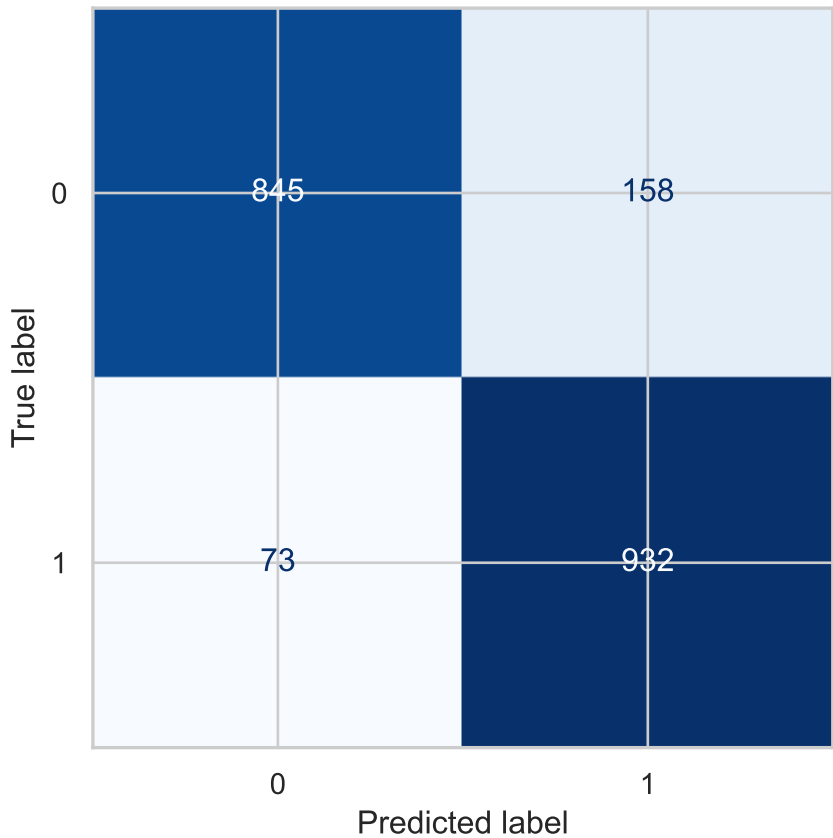
Cluster 6 - top TF-IDF words (centroid)



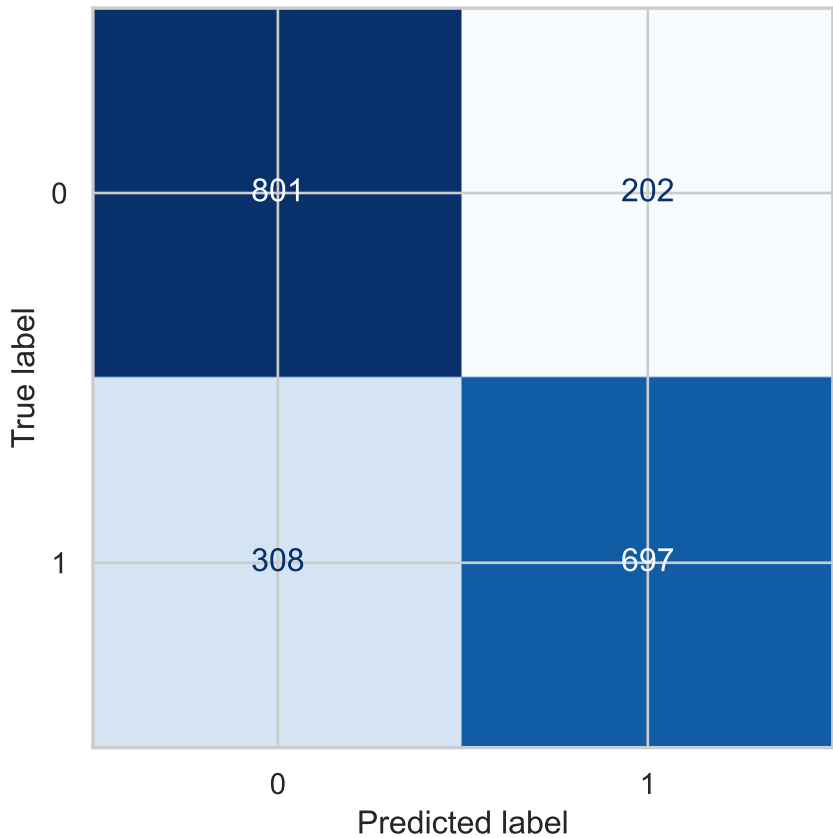
Naive Bayes - Confusion Matrix



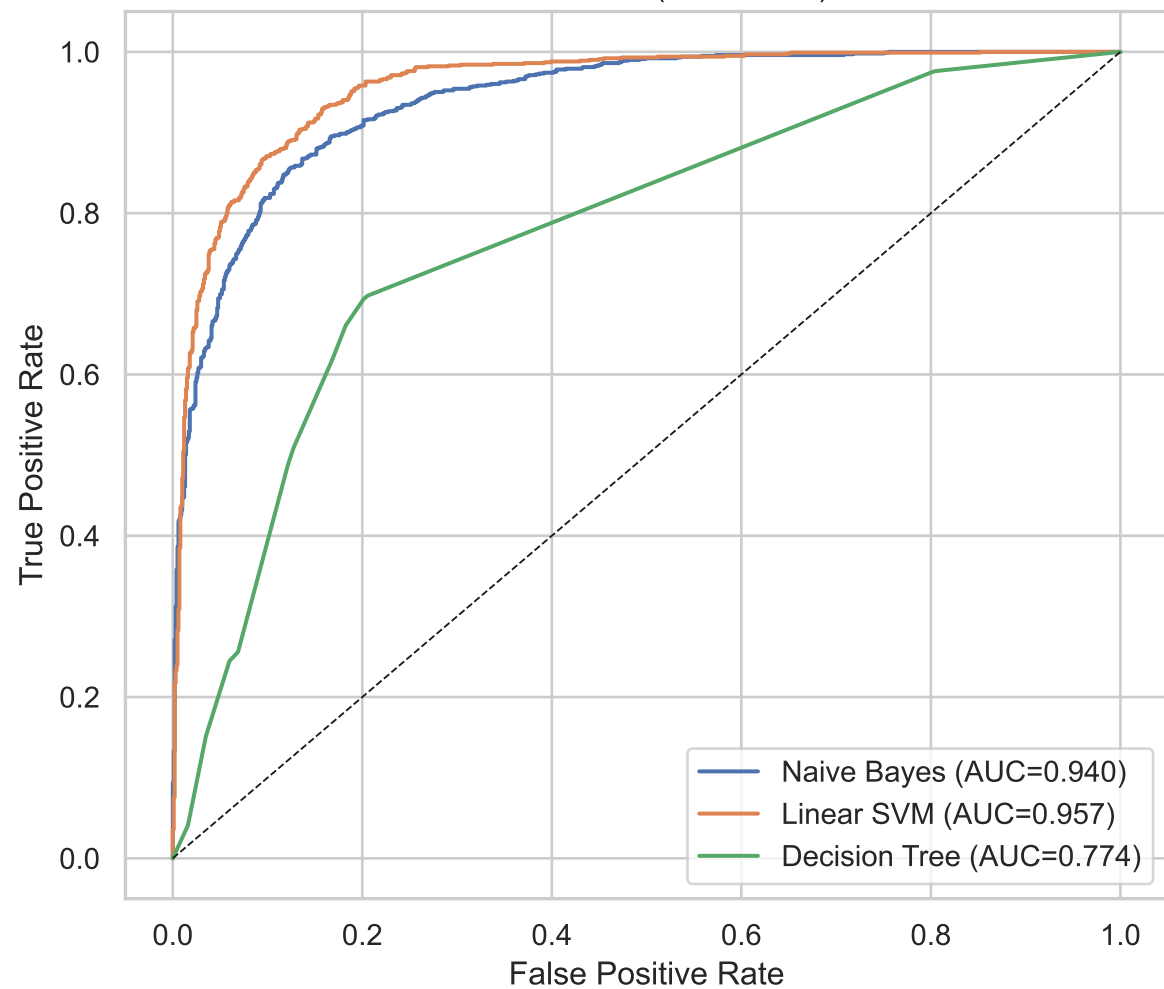
Linear SVM - Confusion Matrix



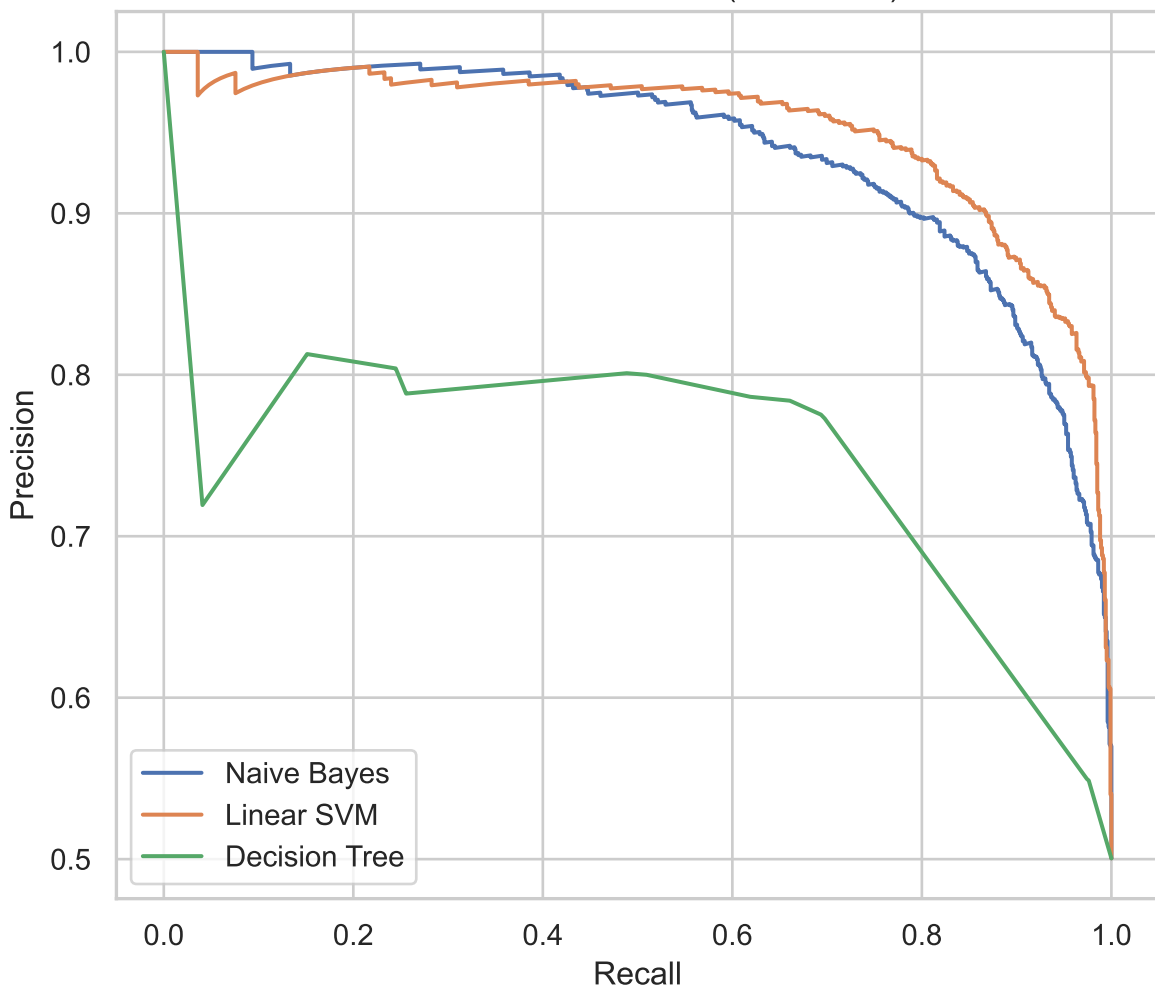
Decision Tree - Confusion Matrix



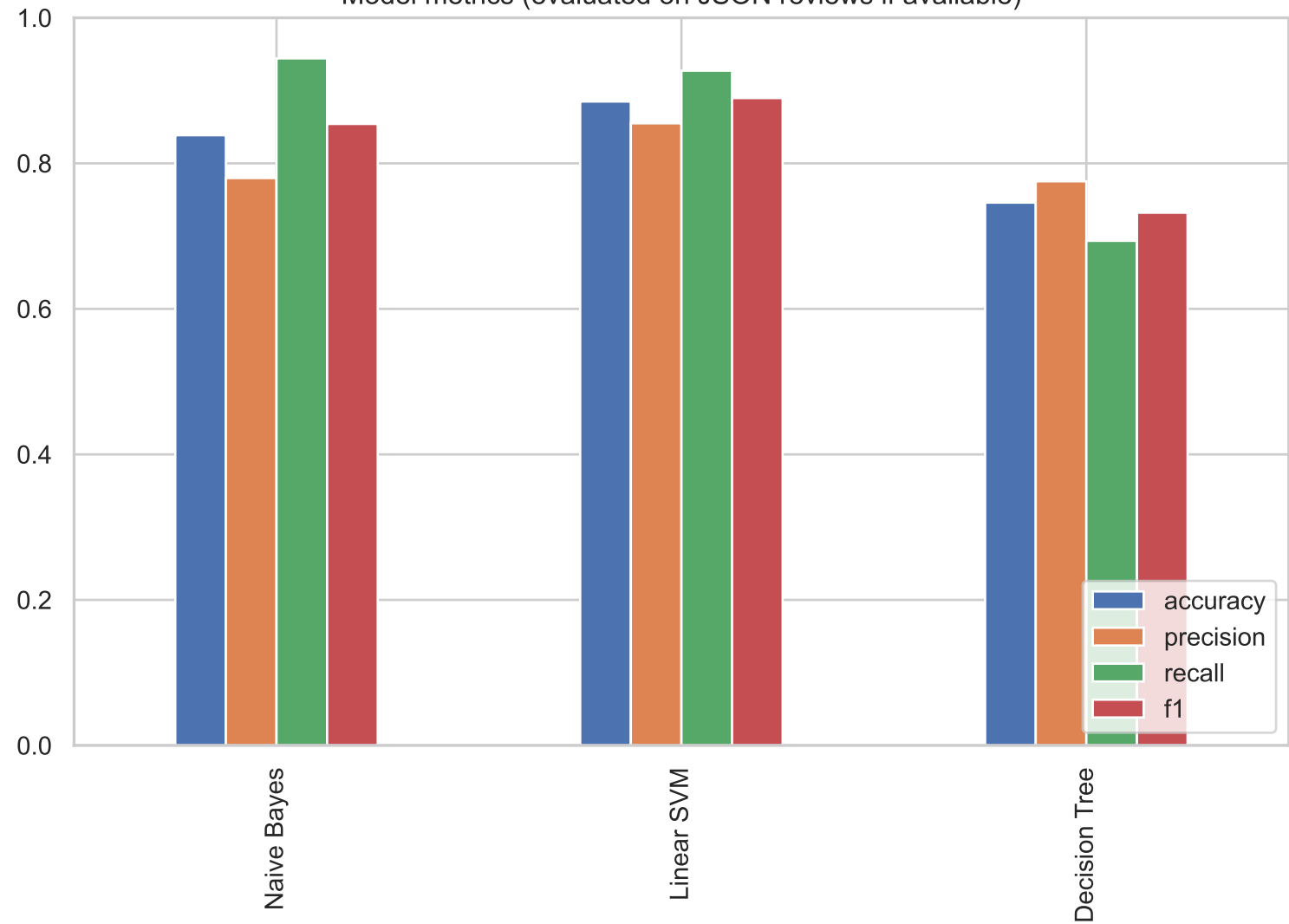
ROC Curves (JSON eval)



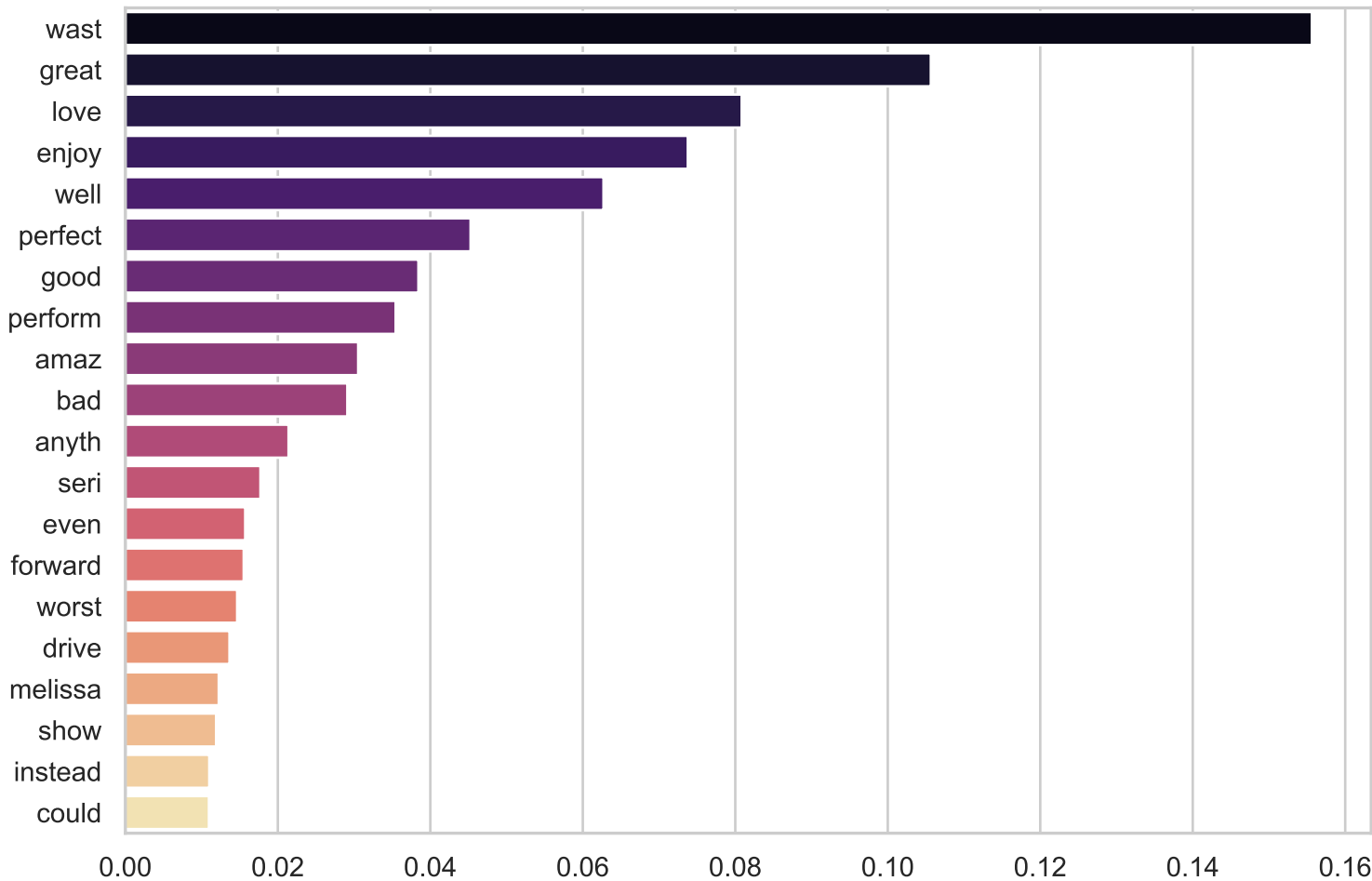
Precision-Recall Curves (JSON eval)



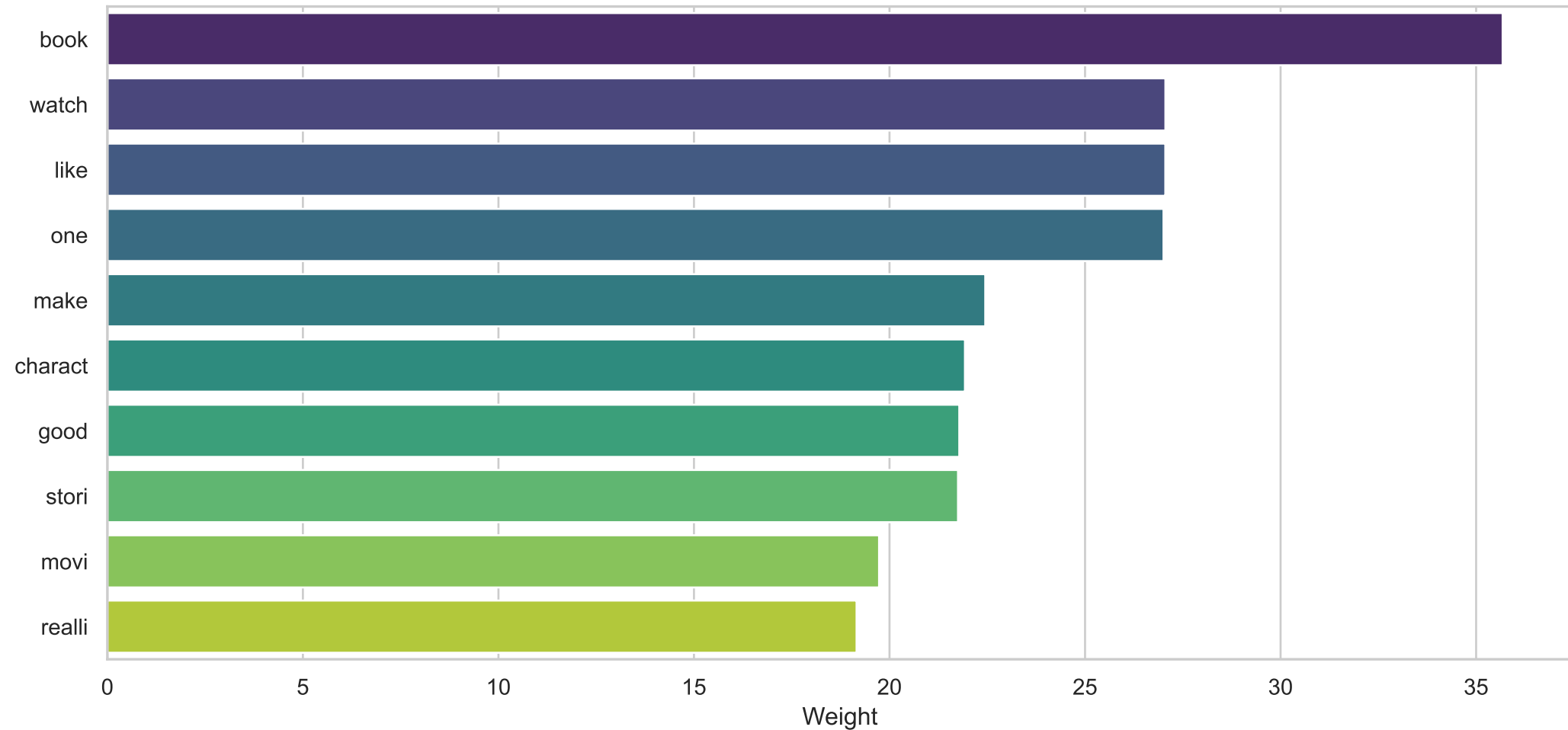
Model metrics (evaluated on JSON reviews if available)



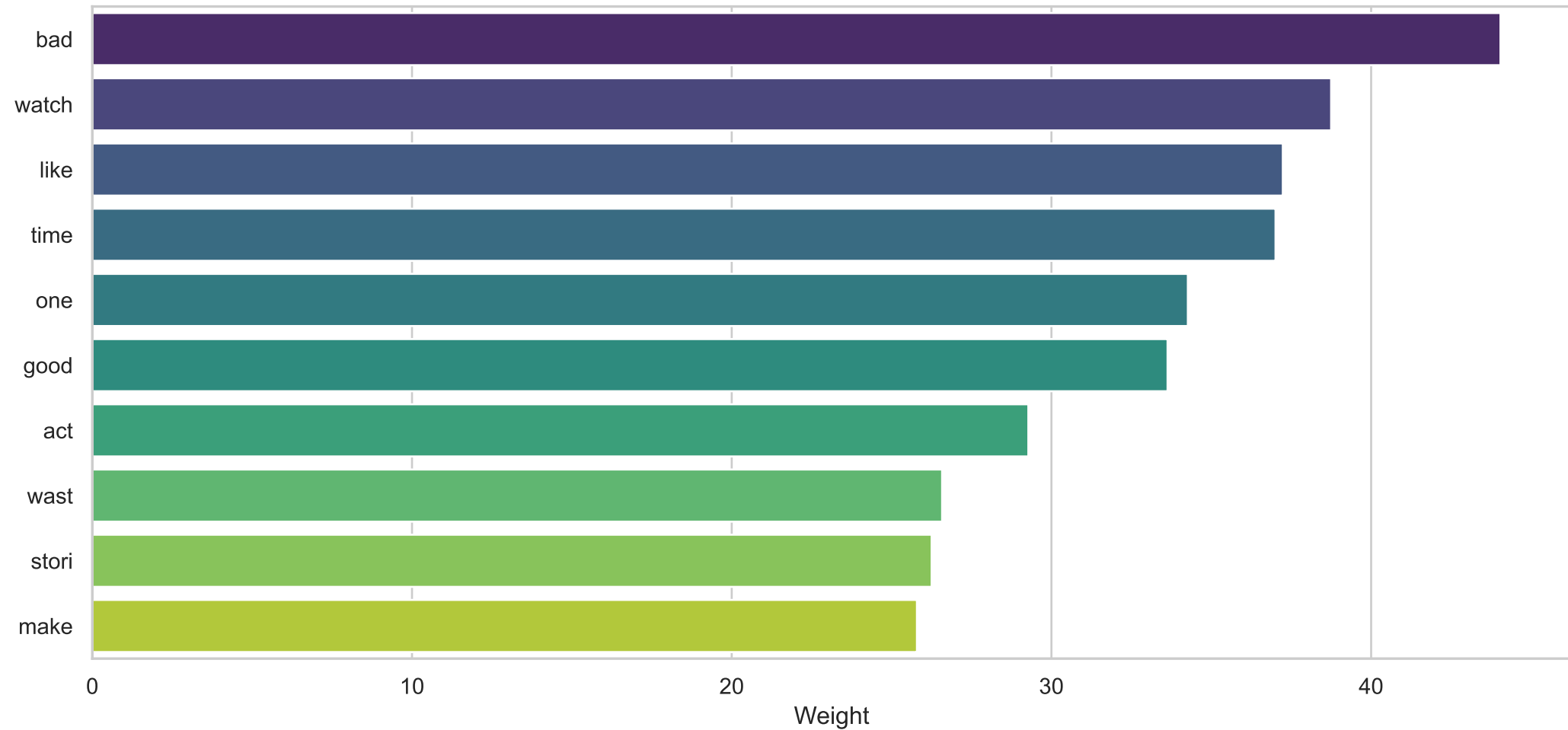
Decision Tree - top feature importances (training)



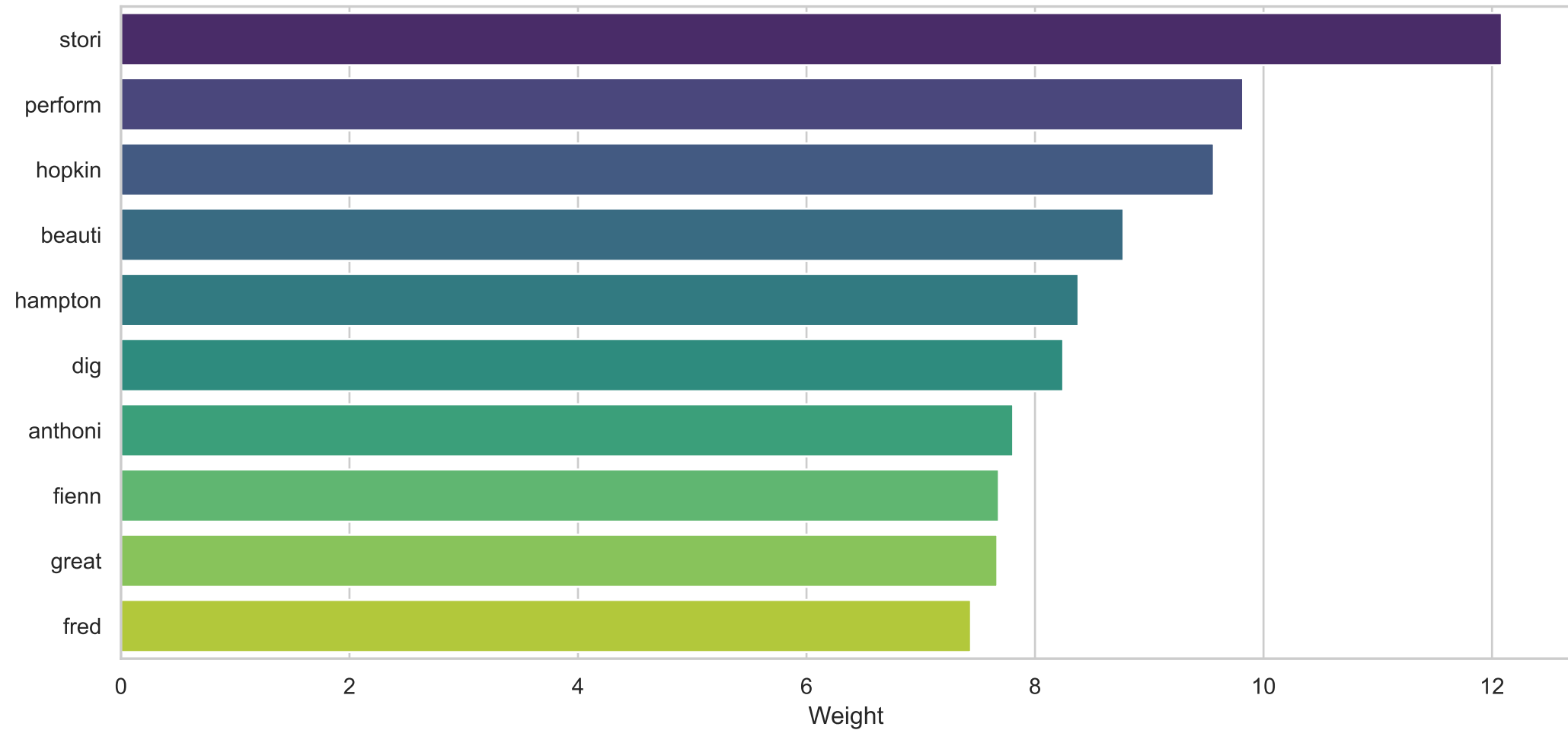
LDA Topic 0 - top words (LDA)



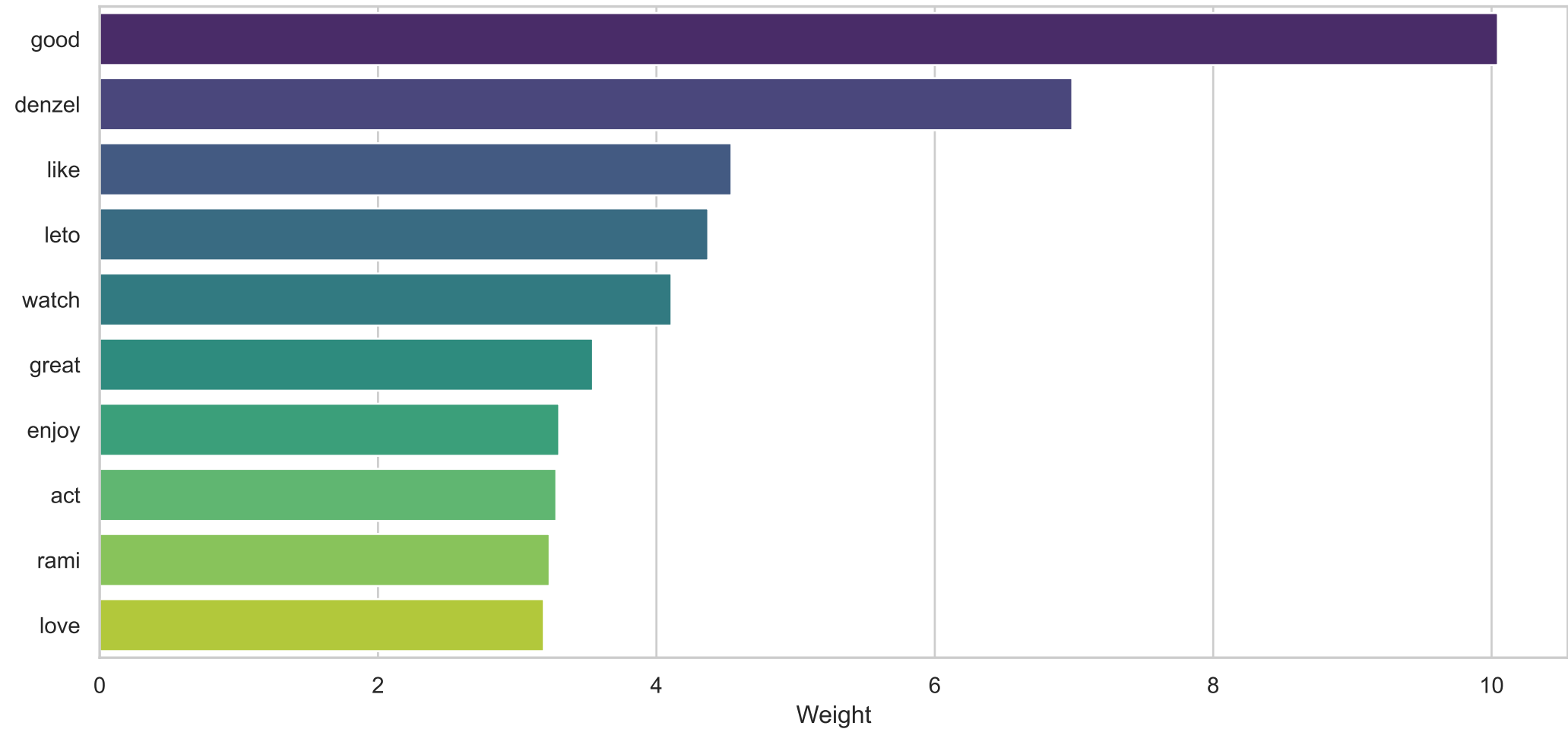
LDA Topic 1 - top words (LDA)



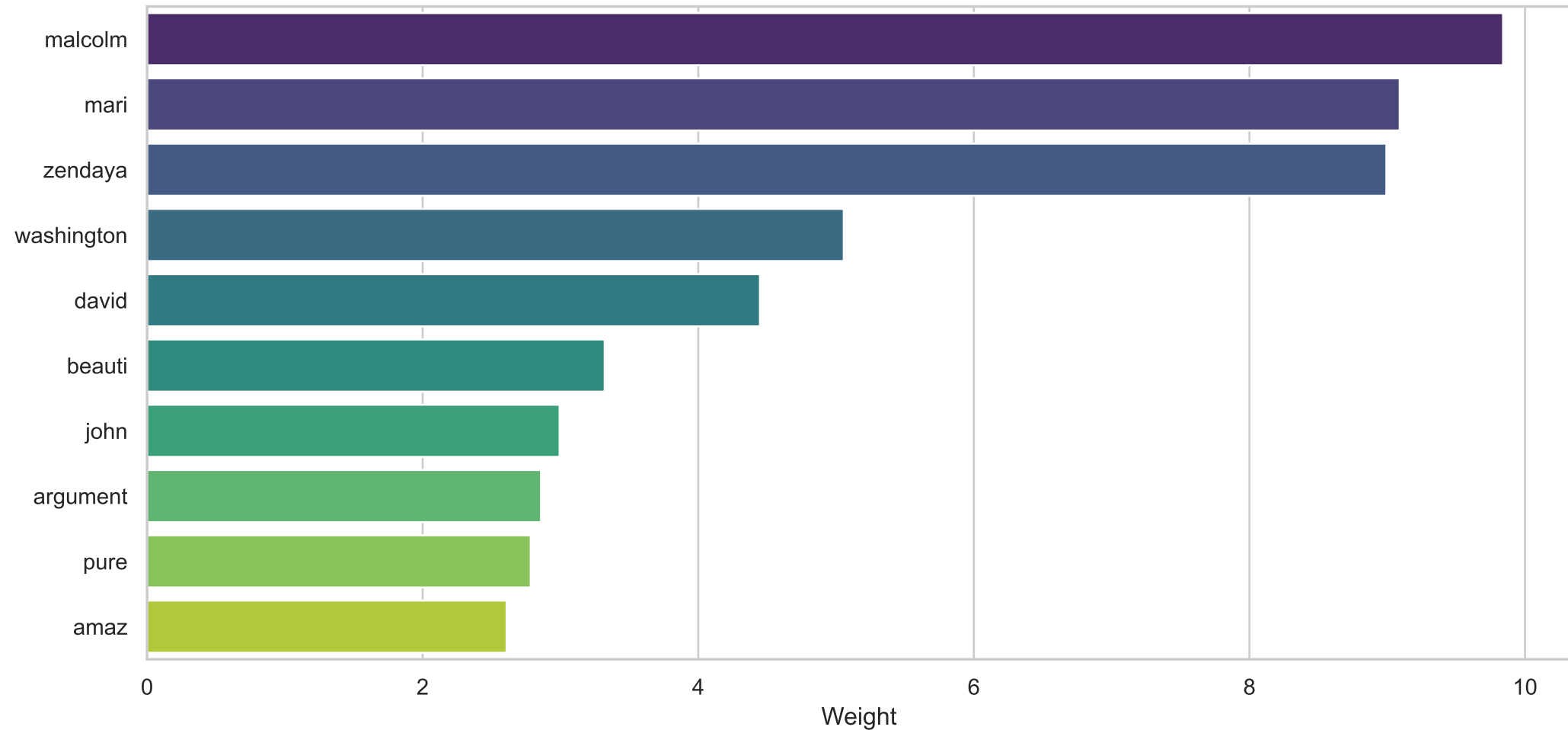
LDA Topic 2 - top words (LDA)



LDA Topic 3 - top words (LDA)



LDA Topic 4 - top words (LDA)



Classification reports (evaluated on JSON reviews if available)

Model: Naive Bayes					
	precision	recall	f1-score	support	
NEG	0.93	0.73	0.82	1003	
POS	0.78	0.94	0.85	1005	
accuracy			0.84	2008	
macro avg	0.85	0.84	0.84	2008	
weighted avg	0.85	0.84	0.84	2008	

Model: Linear SVM					
	precision	recall	f1-score	support	
NEG	0.92	0.84	0.88	1003	
POS	0.86	0.93	0.89	1005	
accuracy			0.88	2008	
macro avg	0.89	0.88	0.88	2008	
weighted avg	0.89	0.88	0.88	2008	

Model: Decision Tree					
	precision	recall	f1-score	support	
NEG	0.72	0.80	0.76	1003	
POS	0.78	0.69	0.73	1005	
accuracy			0.75	2008	
macro avg	0.75	0.75	0.75	2008	
weighted avg	0.75	0.75	0.75	2008	