

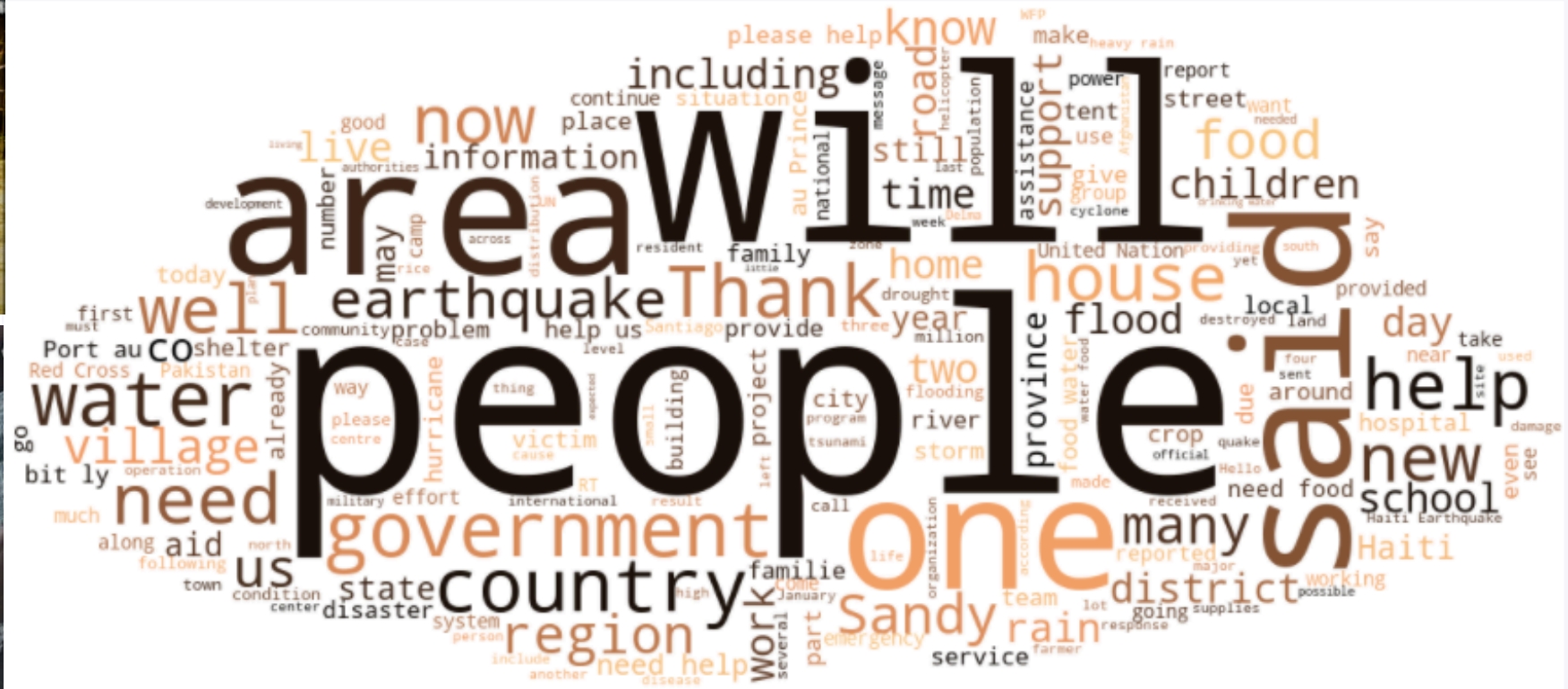


NLP UNSUPERVISED LEARNING: MESSAGE CLASSIFICATION FOR DISASTER RESPONSE

METIS BOOTCAMP | July 23, 2021
Presented by: Kalu Uga

PROBLEM STATEMENT

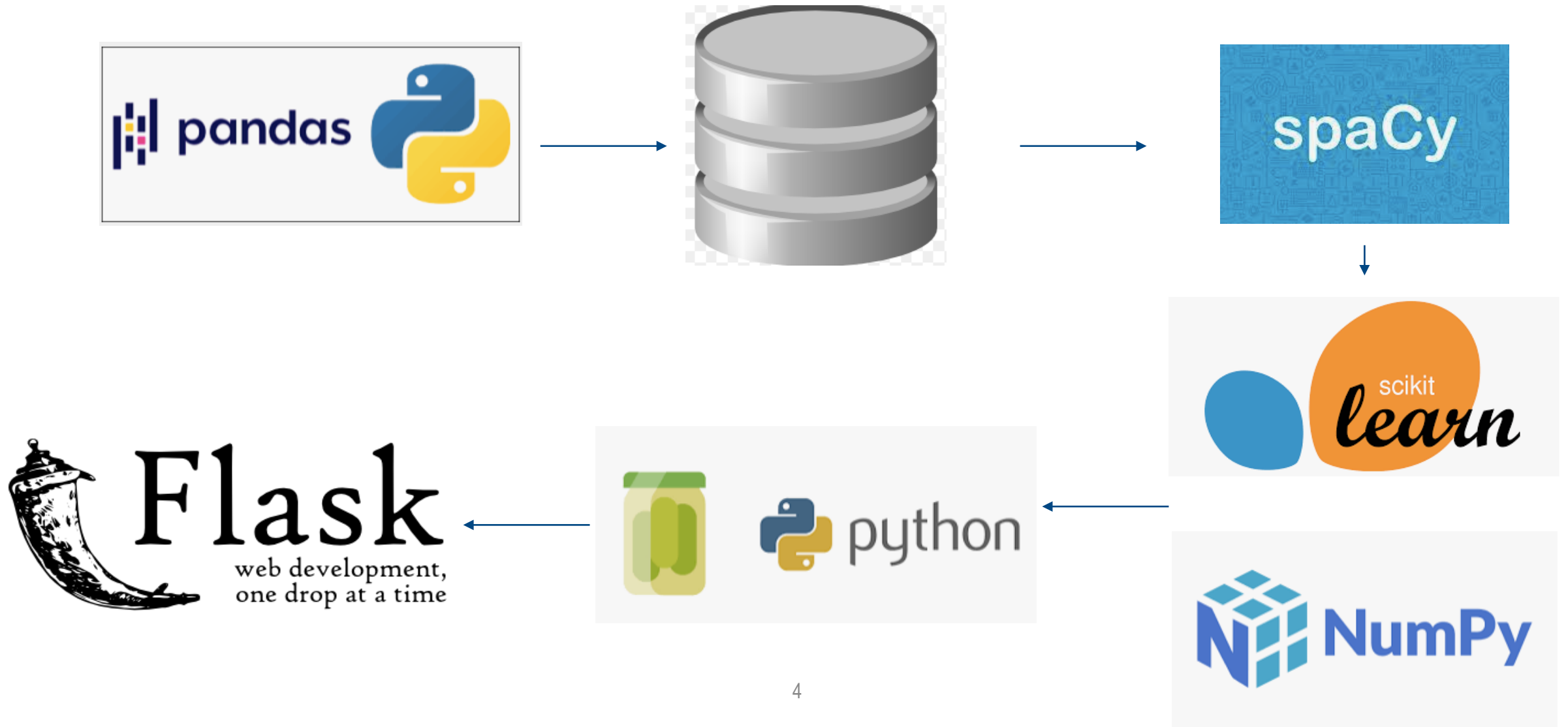
Use topic modeling to group disaster messages and predict the group where a new message belong.



DATASET



WORKFLOW



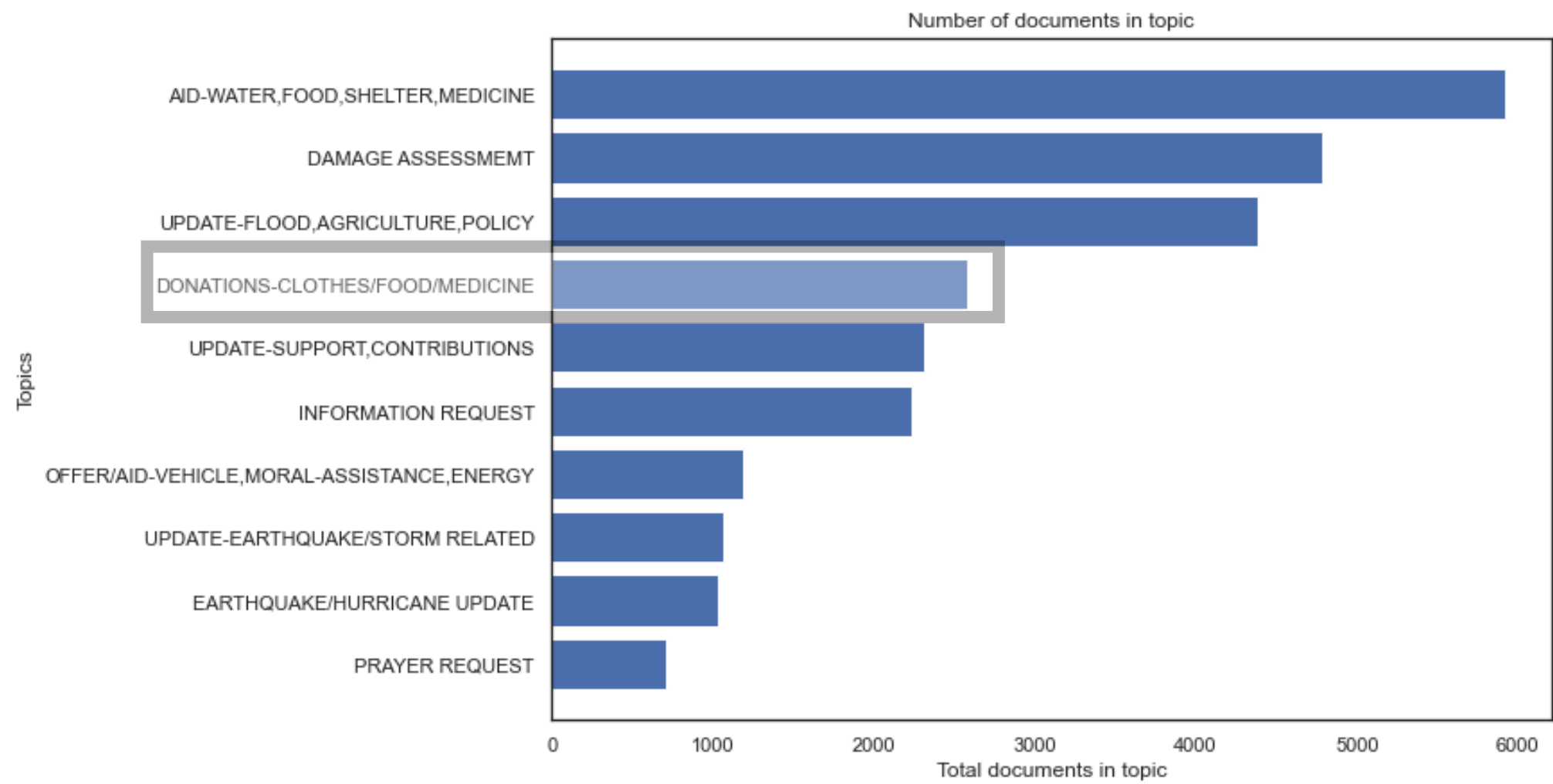
PREPROCESSING

- Clean Documents
- Tokenization
- Lemmatization
- Vectorizer
 - ★ Cut off words that occurs in less than 10 documents
 - ★ Cut off words that occur in more than 50% of document

MODELS

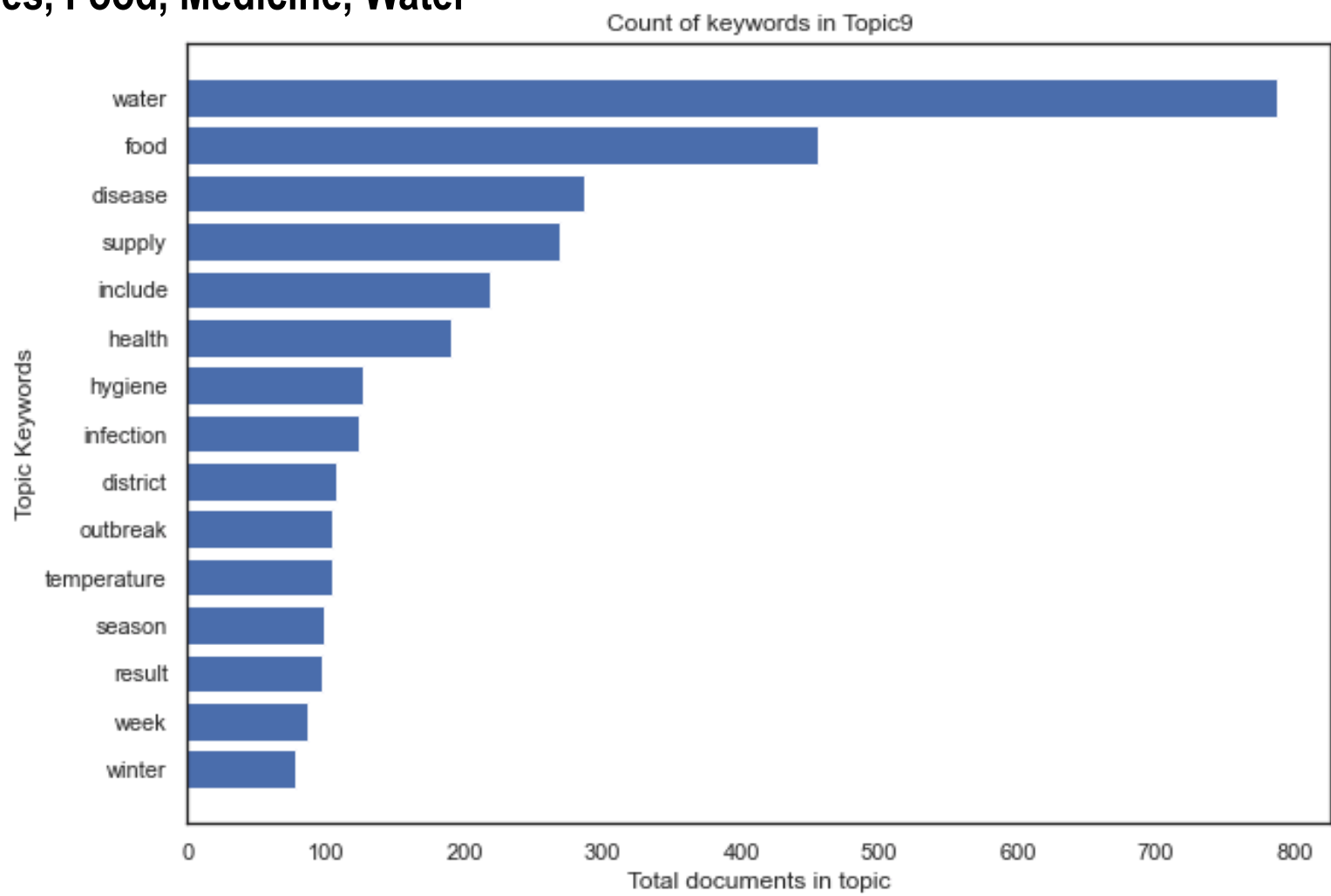
- CountVectorizer + LatentDirichletAllocation
- **TfidfVectorizer + LatentDirichletAllocation - *Best topic interpretation***
- CountVectorizer + TruncatedSVD
- TfidfVectorizer + TruncatedSVD
- CountVectorizer + NMF
- TfidfVectorizer + NMF

TFIDF + LDA - TOPICS (10 TOPICS, 15 KEYWORDS)

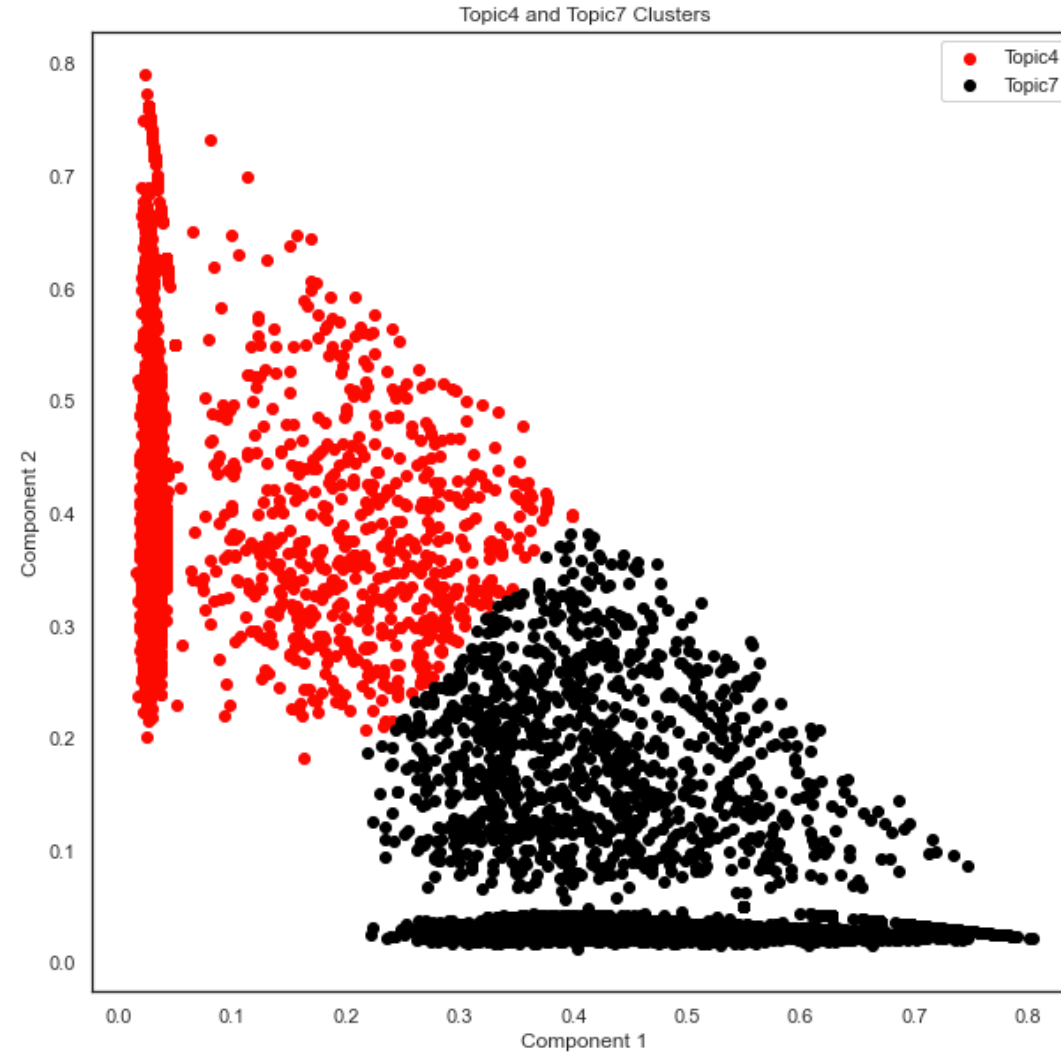


TFIDF + LDA - KEYWORDS IN TOPIC9

★ Donations - Clothes, Food, Medicine, Water



FILTERED CLUSTER - TOPIC4 AND TOPIC7



PREDICTION AND MOST SIMILAR MESSAGE

- Message Example

- ★ ["We need food and medical supplies. The kids also need blankets and clothing"]

- Predicted group

- ★ Aid -Water, Food, Shelter, Medicine

- Most Similar Message - Cosine Similarity

- ★ we have clothes for all ages and brand new shoes and boots we have baby items like diapers car seats high chair ect we also have hygiene products like wipes toothbrushes toothpaste ect we also have some kids toys and stuffed animals even a bike and we also have some non perishable foods and blankets

PREDICTION - FLASK APP

Topic Modeling App

Disaster Message Categorization

Please Enter Message Below

We need food and medical supplies. The kids also need blankets and clothing

Get Category

Topic Modeling App

Disaster Message Categorization

Message Category

AID-WATER,FOOD,SHELTER,MEDICINE

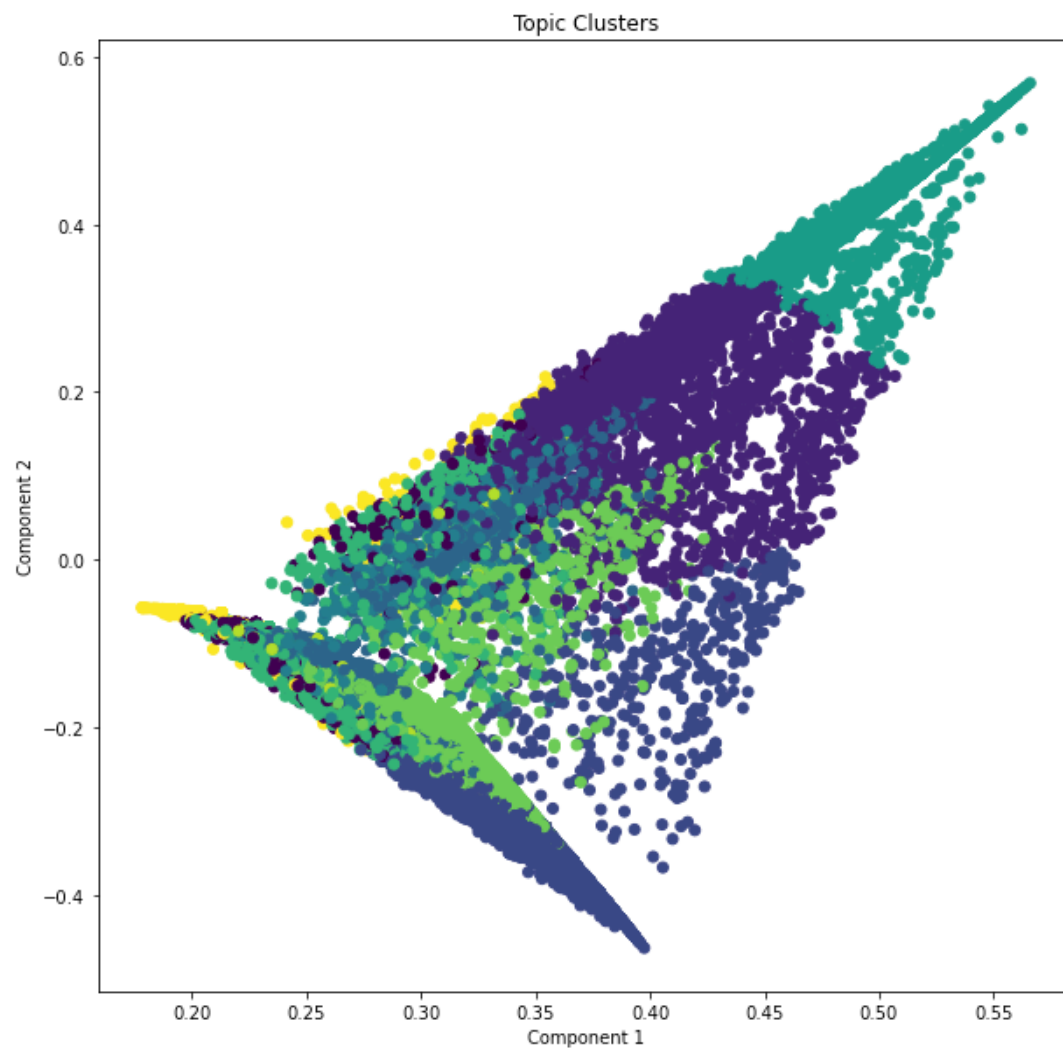
RECOMMENDATIONS

- Tfidf + LDA provides clear interpretation of topics for a mixture of short and long messages.
- Tfidf + LDA can find topics that other algorithms will miss.

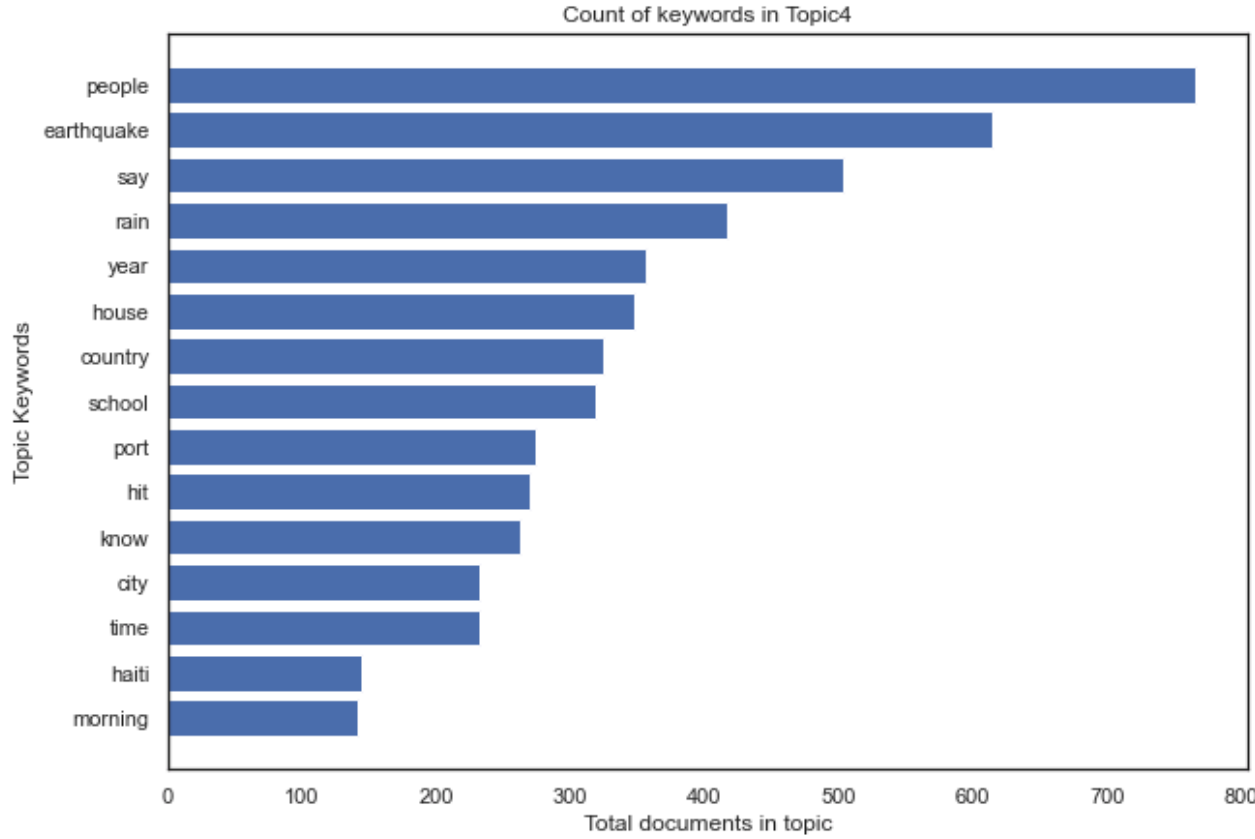
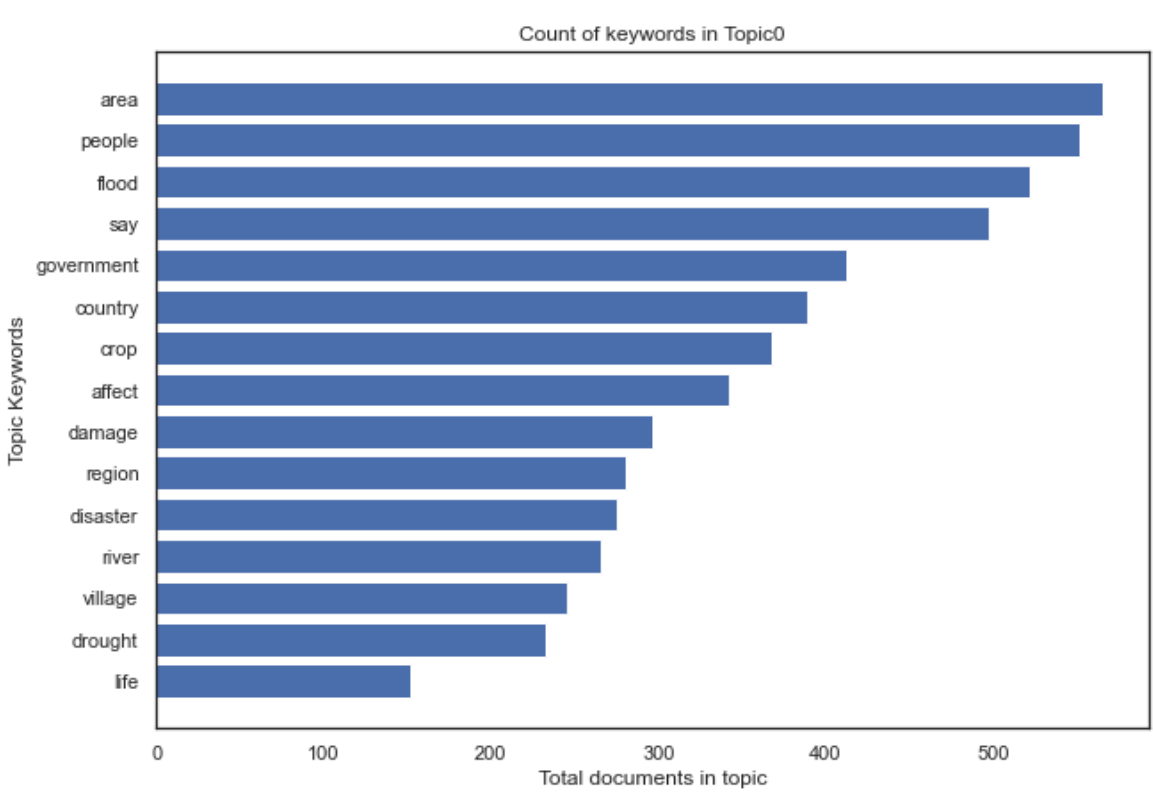
APPENDIX - TFIDF + LDA - TOPICS

- Using 10 topics and 15 keywords
 - ★ Update - Flood, Agriculture
 - ★ Update - Support, Contributions
 - ★ **Aid - Water, Food, Shelter, Medicine**
 - ★ Prayer Request
 - ★ Damage Assessment
 - ★ Update - Earthquake, Storm Related
 - ★ Offer/Aid - Vehicle, Moral Assistance, Energy
 - ★ Information Request
 - ★ Earthquake/Hurricane Update
 - ★ **Donations - Clothes, Food, Medicine, Water**

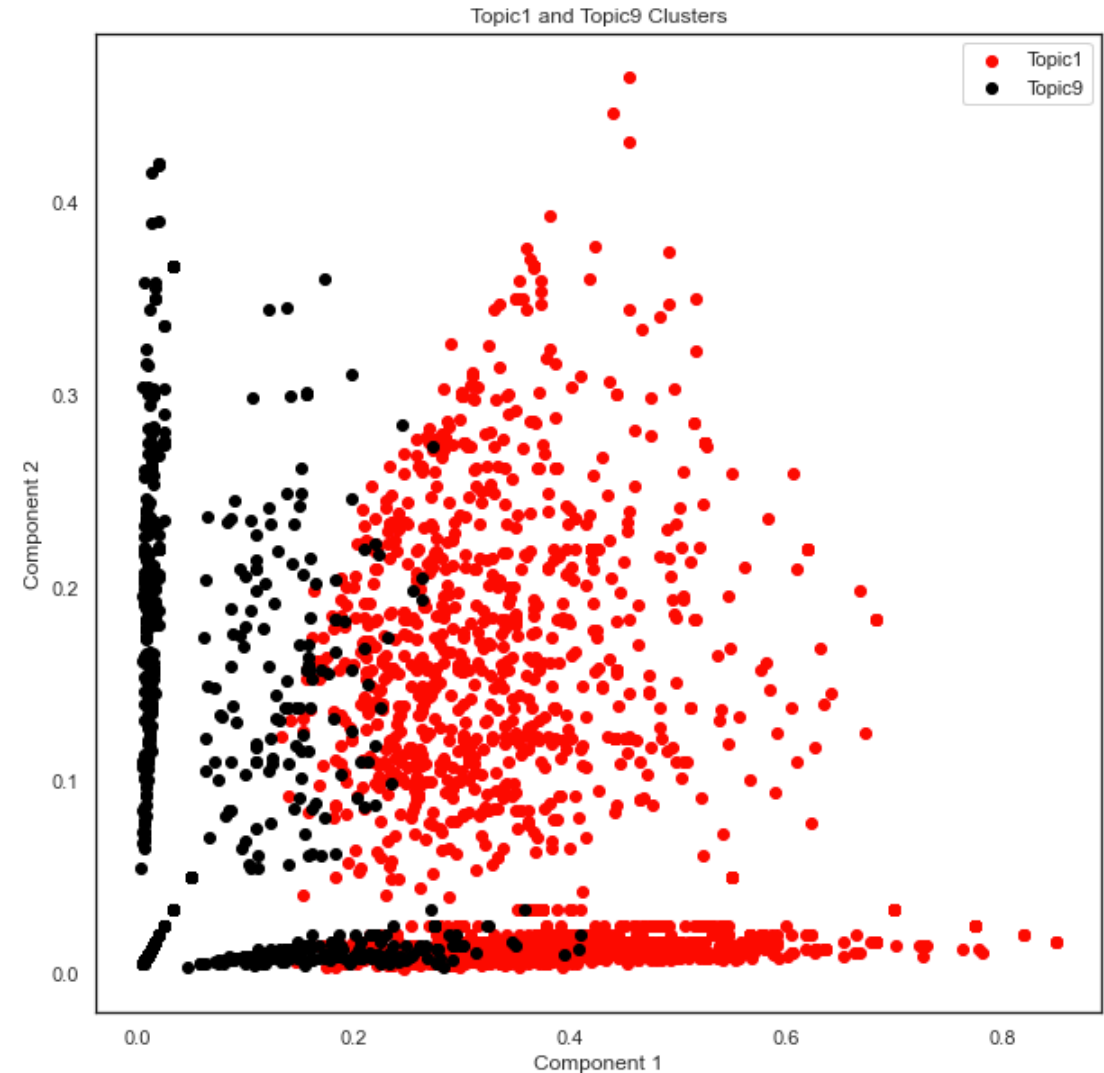
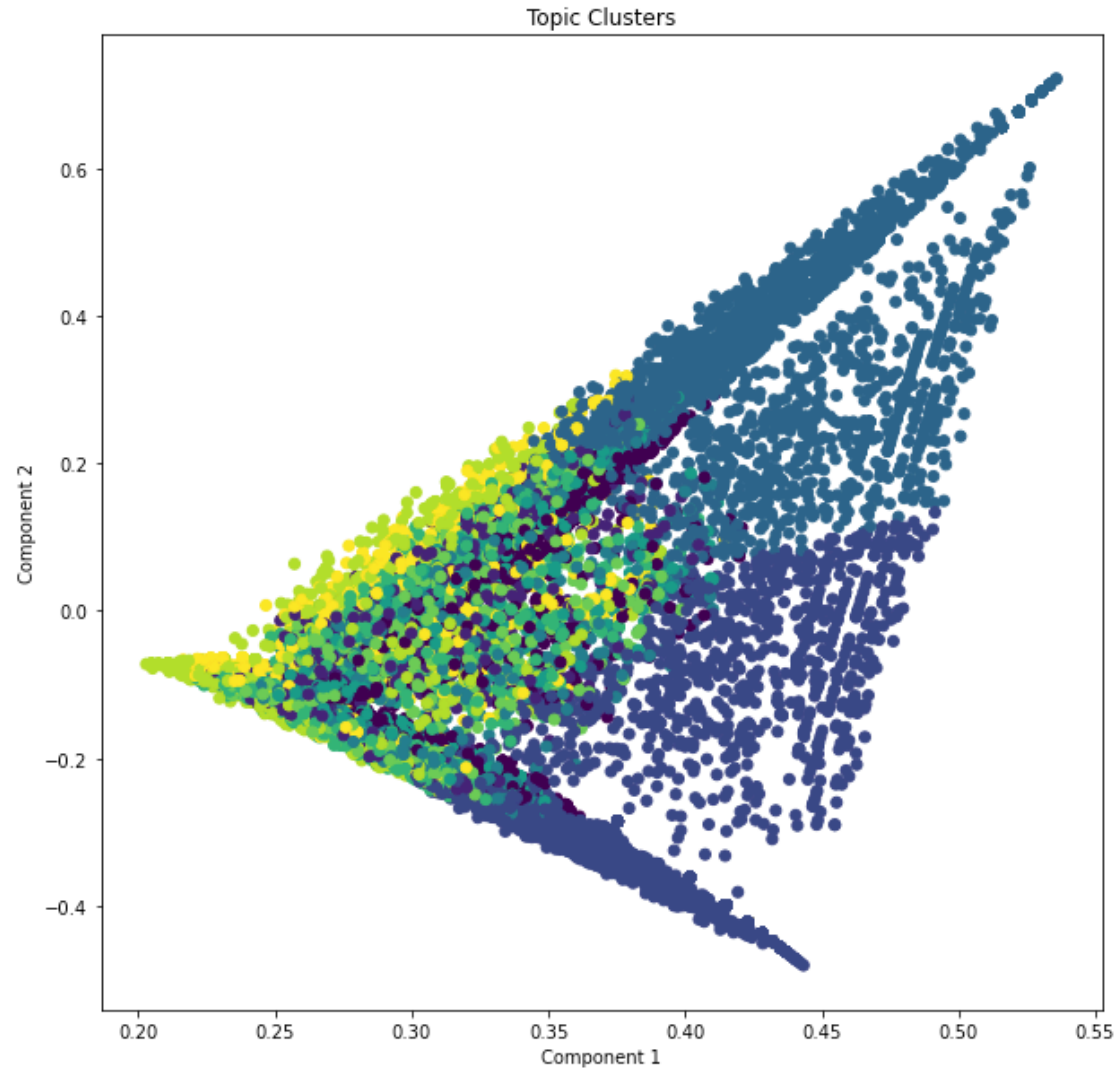
APPENDIX - TFIDF + LDA - CLUSTERS



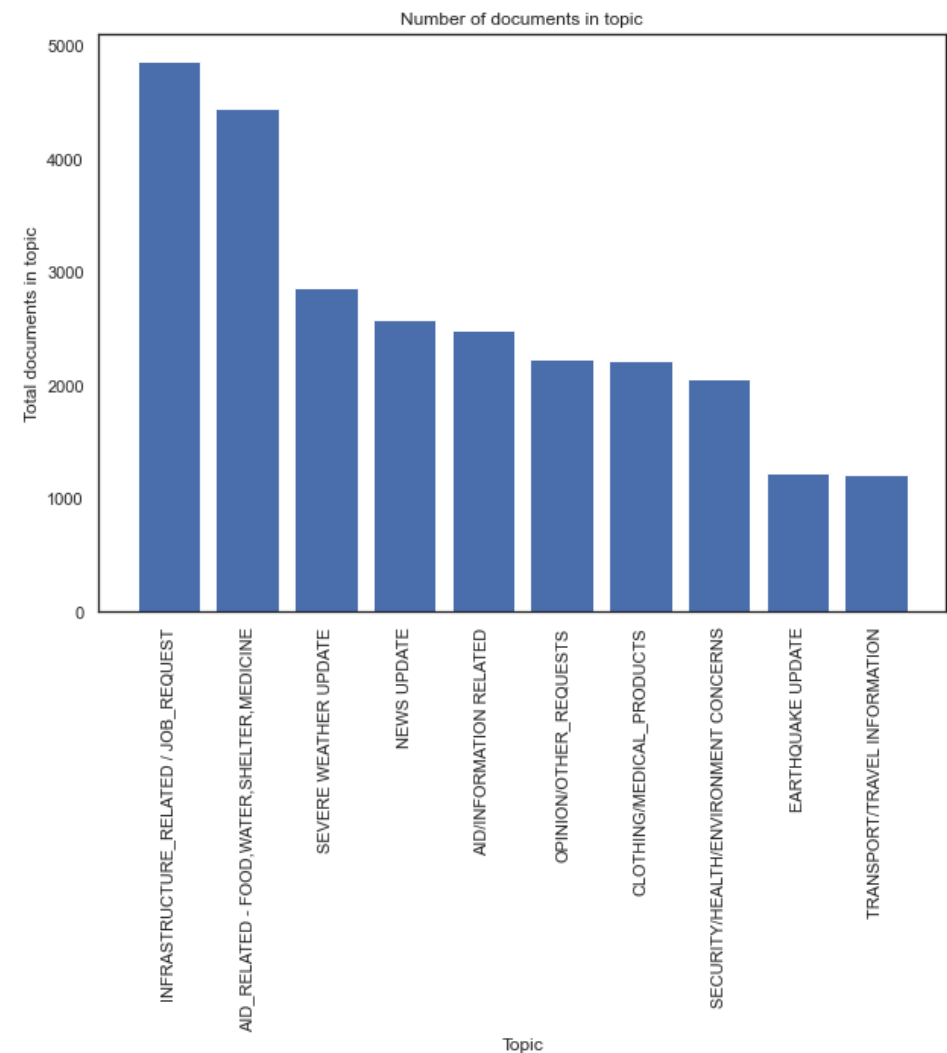
APPENDIX - TFIDF + LDA - TOPIC0/TOPIC4



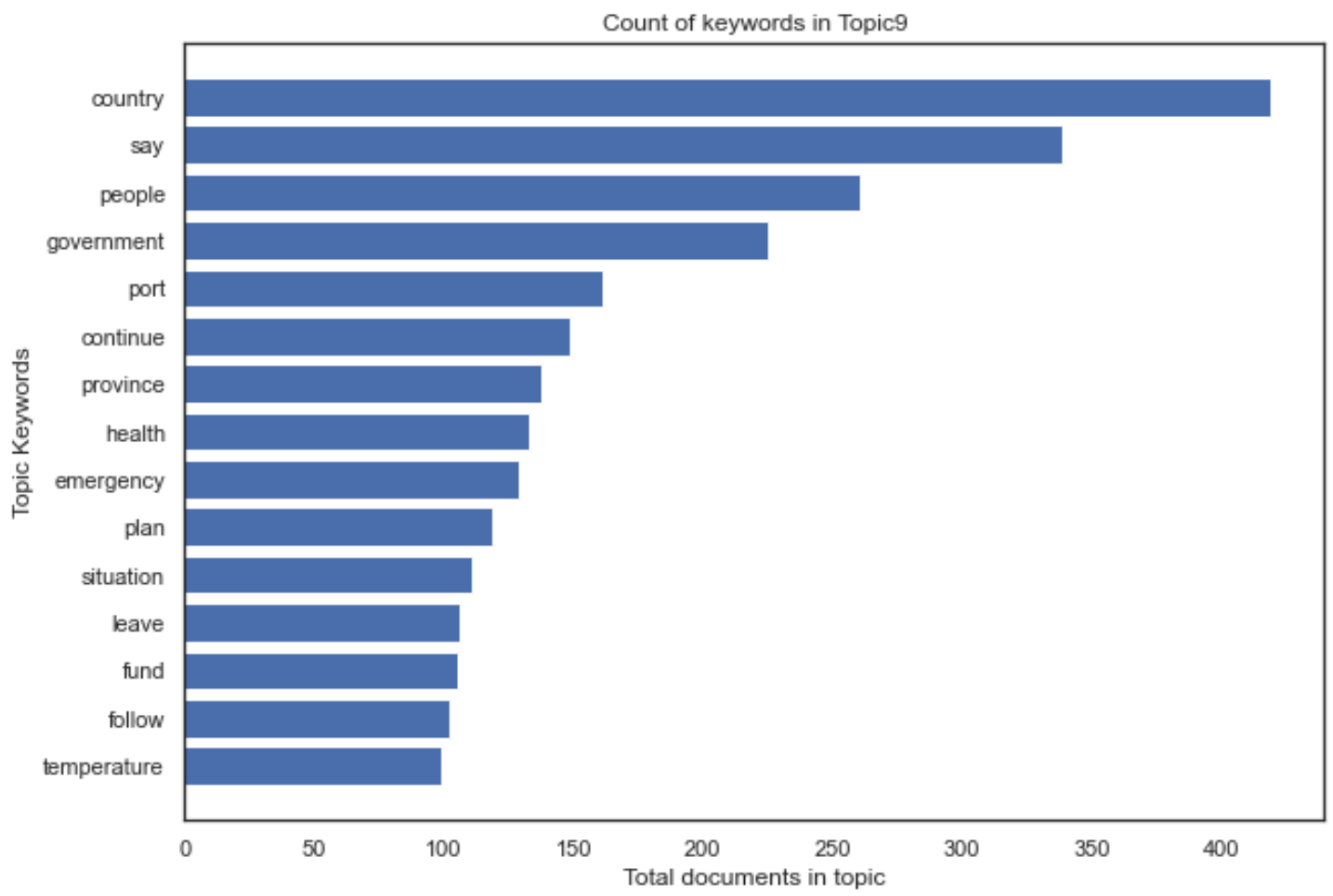
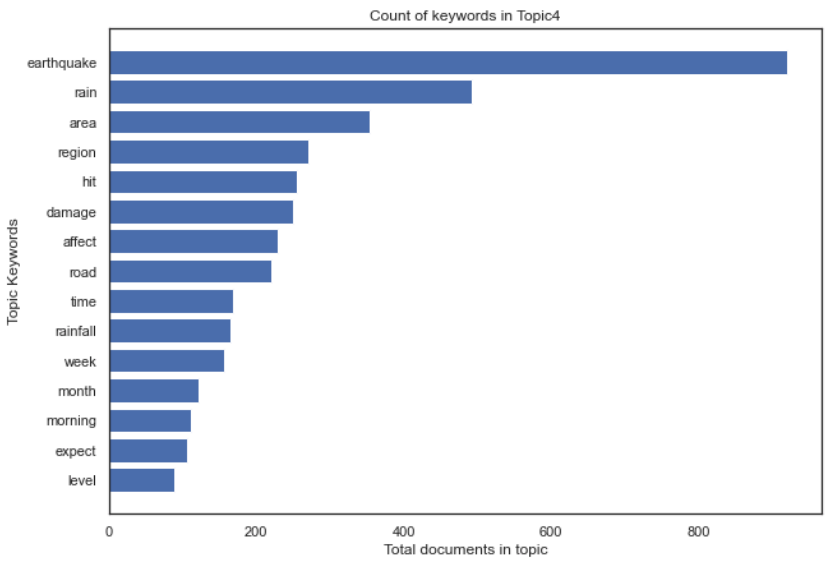
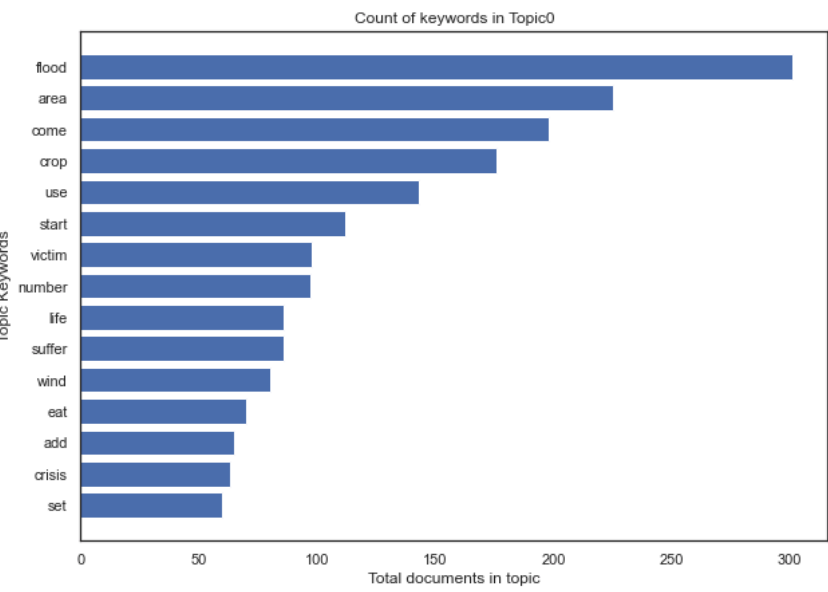
APPENDIX - CV + LDA - CLUSTERS



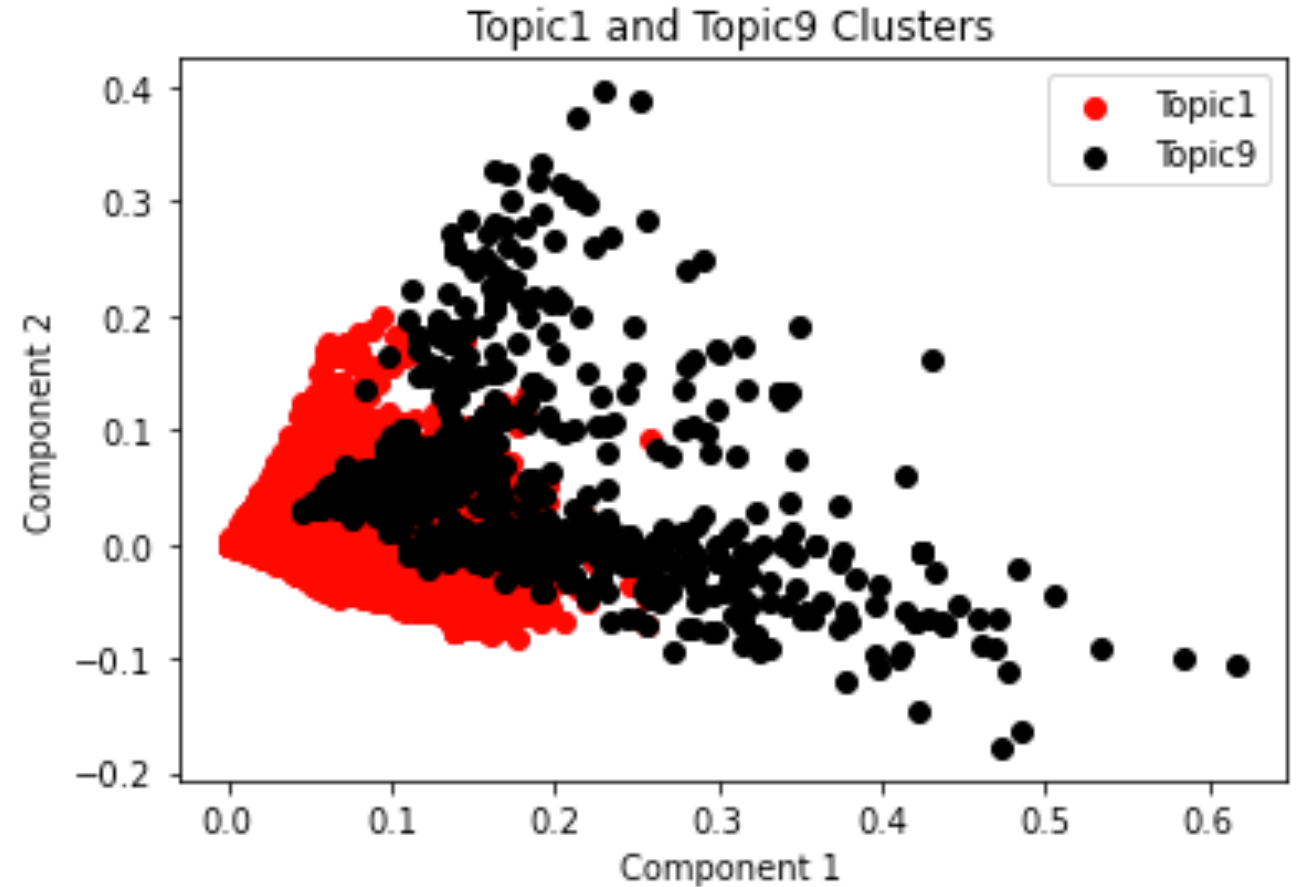
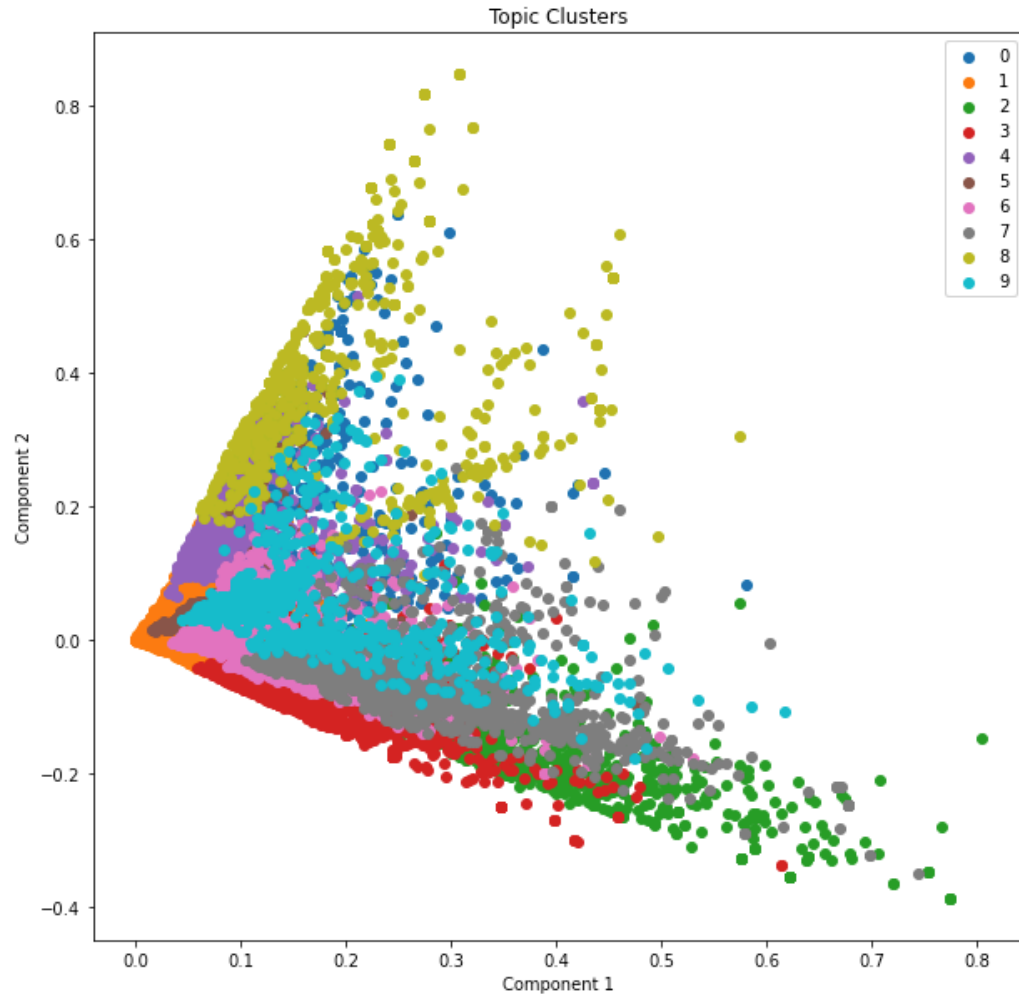
APPENDIX - CV + LDA - TOPIC/DOC COUNT



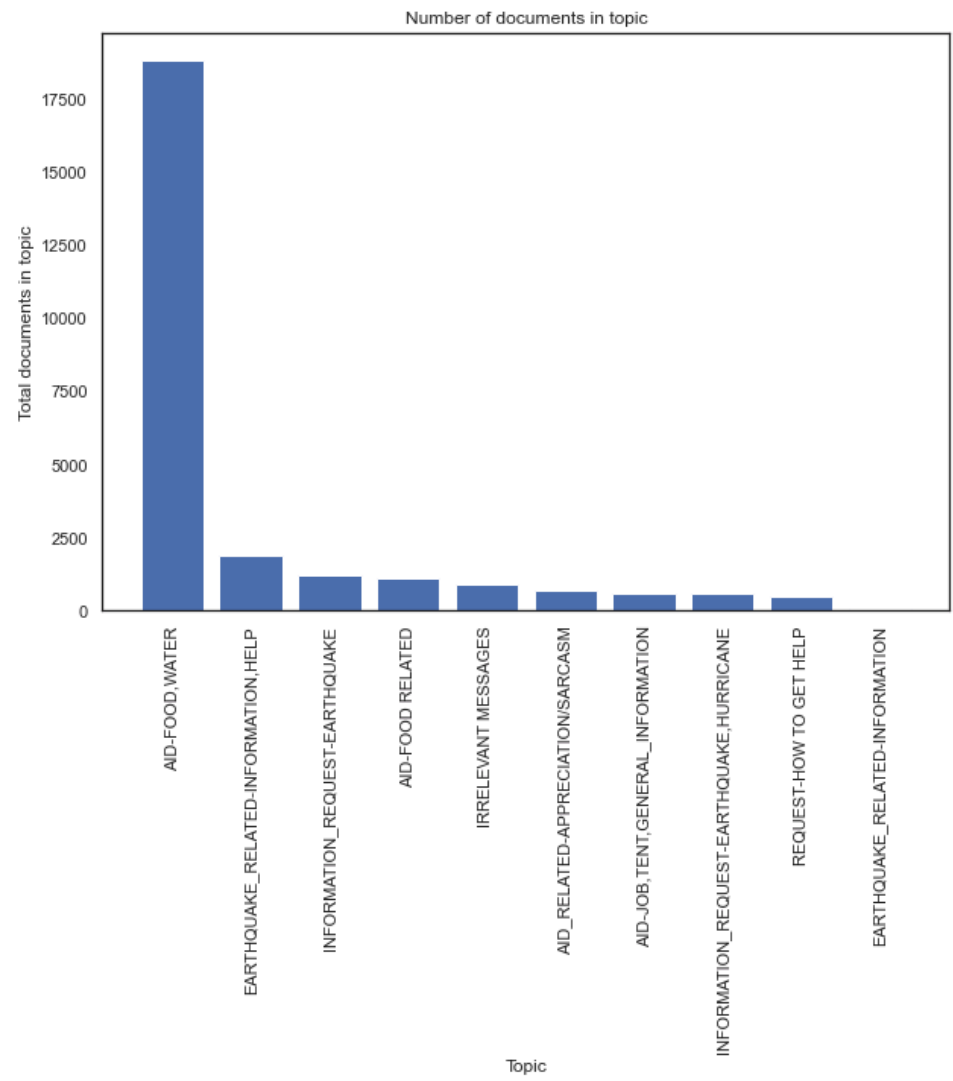
APPENDIX - CV + LDA - TOPIC0/TOPIC4/TOPIC9



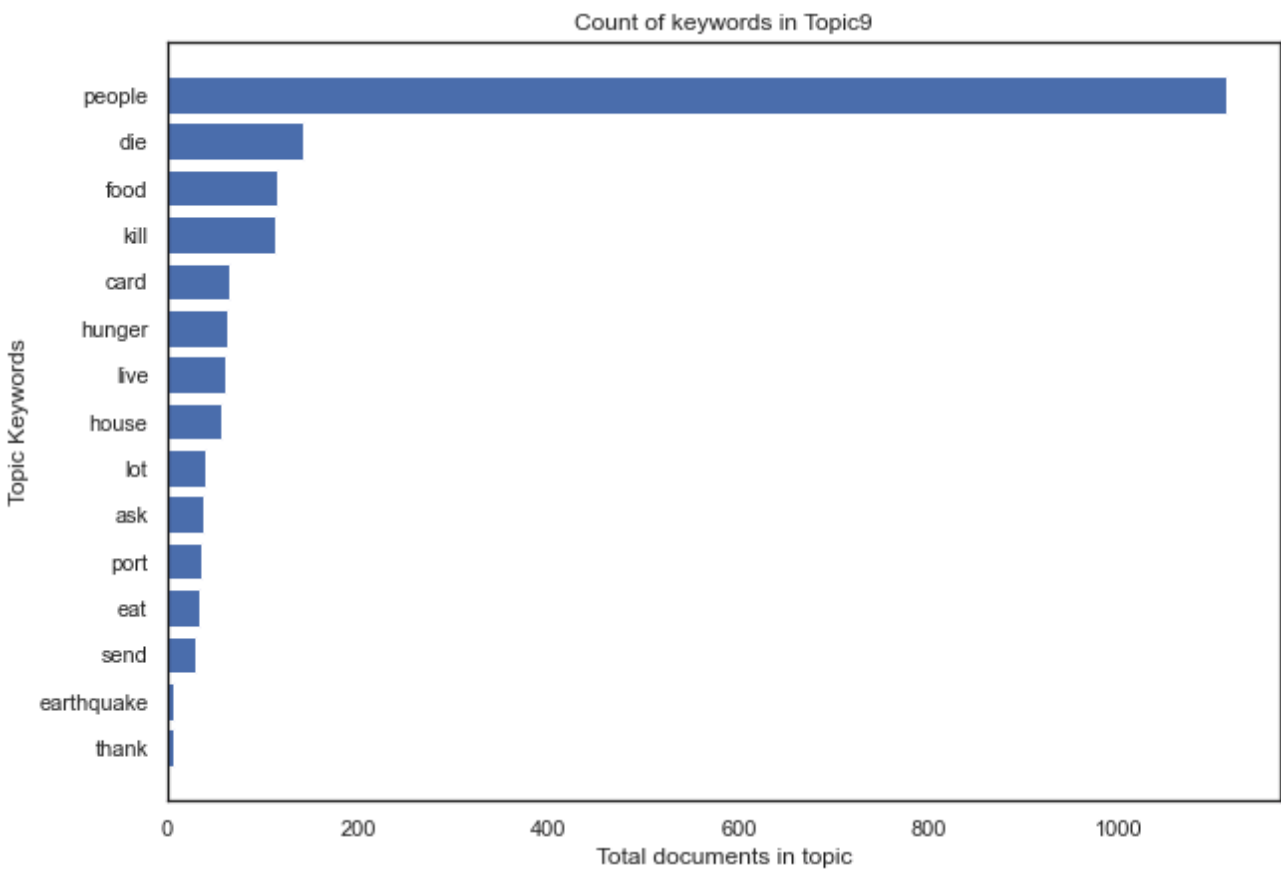
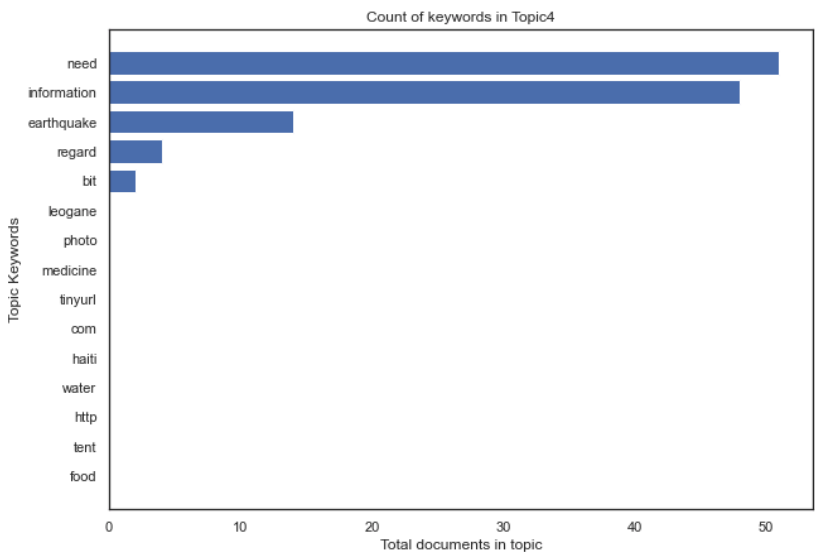
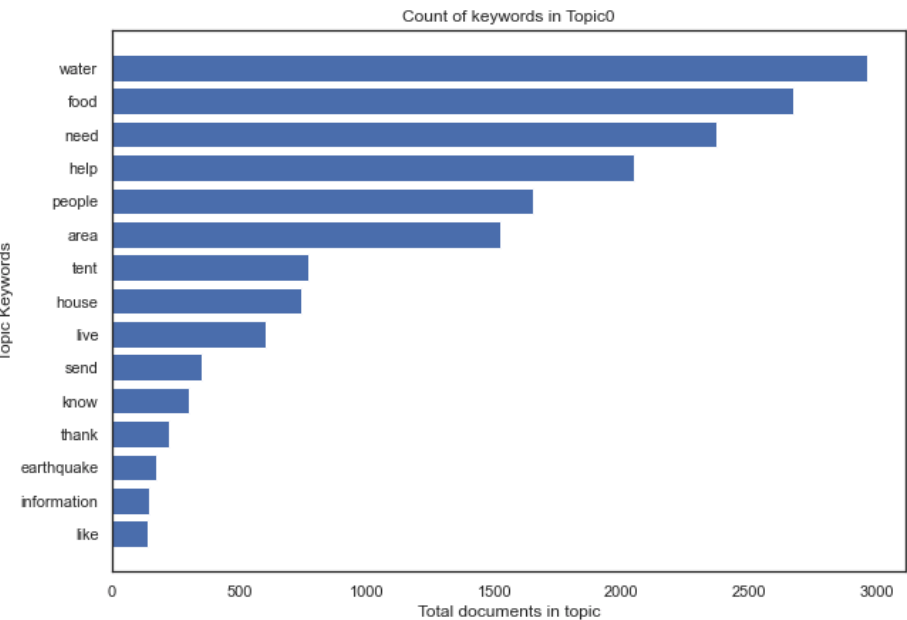
APPENDIX - TFIDF + LSA - CLUSTERS



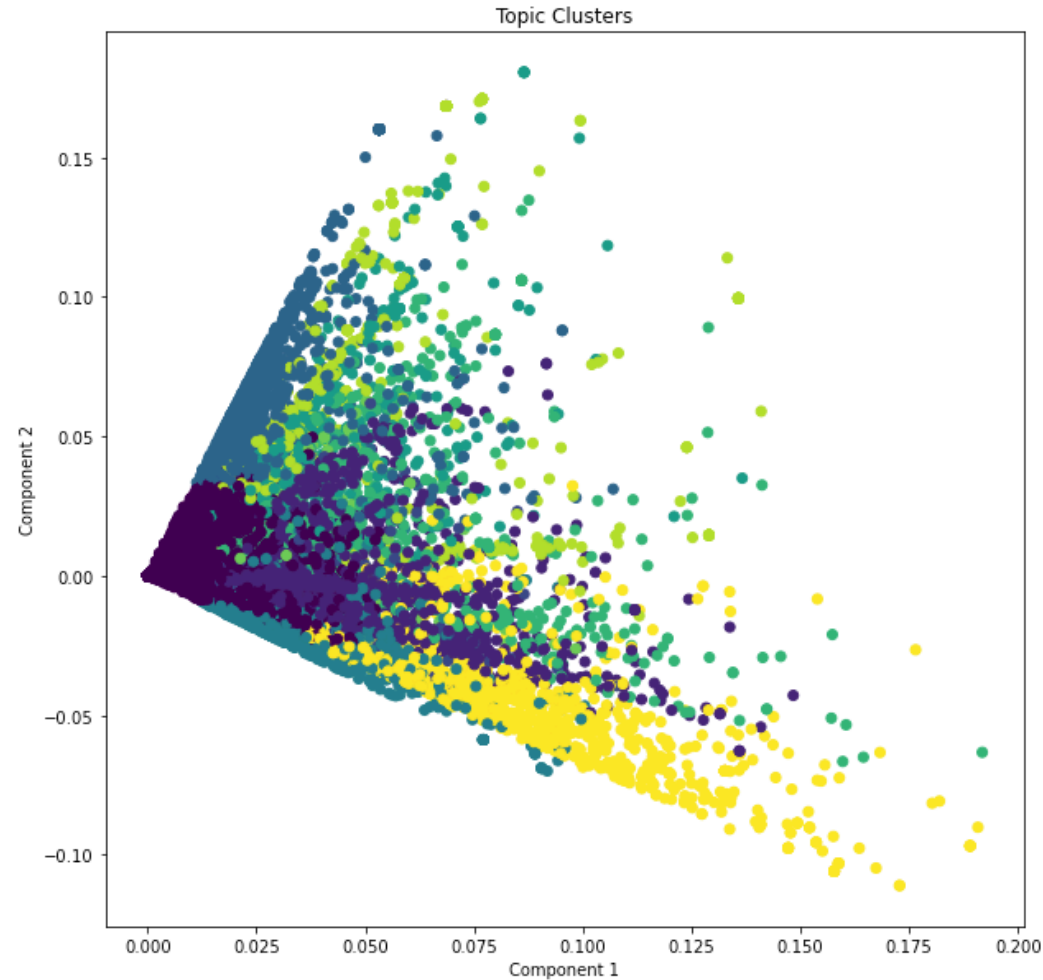
APPENDIX - TFIDF + LSA - TOPIC/DOC COUNT



APPENDIX - TFIDF + LSA - TOPIC0/TOPIC4/TOPIC9



APPENDIX - TFIDF + NMF - CLUSTERS



APPENDIX - TFIDF + NMF - TOPIC4/TOPIC9

