COVID - 19

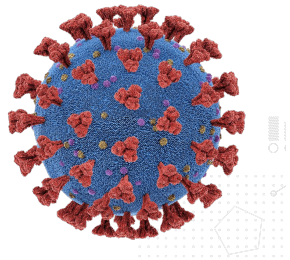
A CASE-STUDY

CONTENTS

COVID-19 THROUGH DATA LENS

INSIGHTS
- Gain better understanding of the pandemic - trends, demographics

FUTURE
- Forecasting new cases for the near future



RESEARCH
- Bring together multiple data sources

SOCIAL PULSE
- See how communities lift themselves up in crisis

IMPACT
- Monitor impact in the areas of global health, government responses & economy

THE JOURNEY

01

RESEARCH & INSIGHTS

CURRENT SITUATION

1. Global Health Indicators
2. Demographic Indicators
3. Government Response Indicators
4. Economic Indicators

Viz tool

Tableau

02

SOCIAL INTERACTIONS

SOCIAL INDICATORS

Understanding social media interactions through topic modeling of tweets relating to Covid-19

Models

Latent Dirichlet Allocation

Non-Negative Matrix Factorization

03

FORECASTING

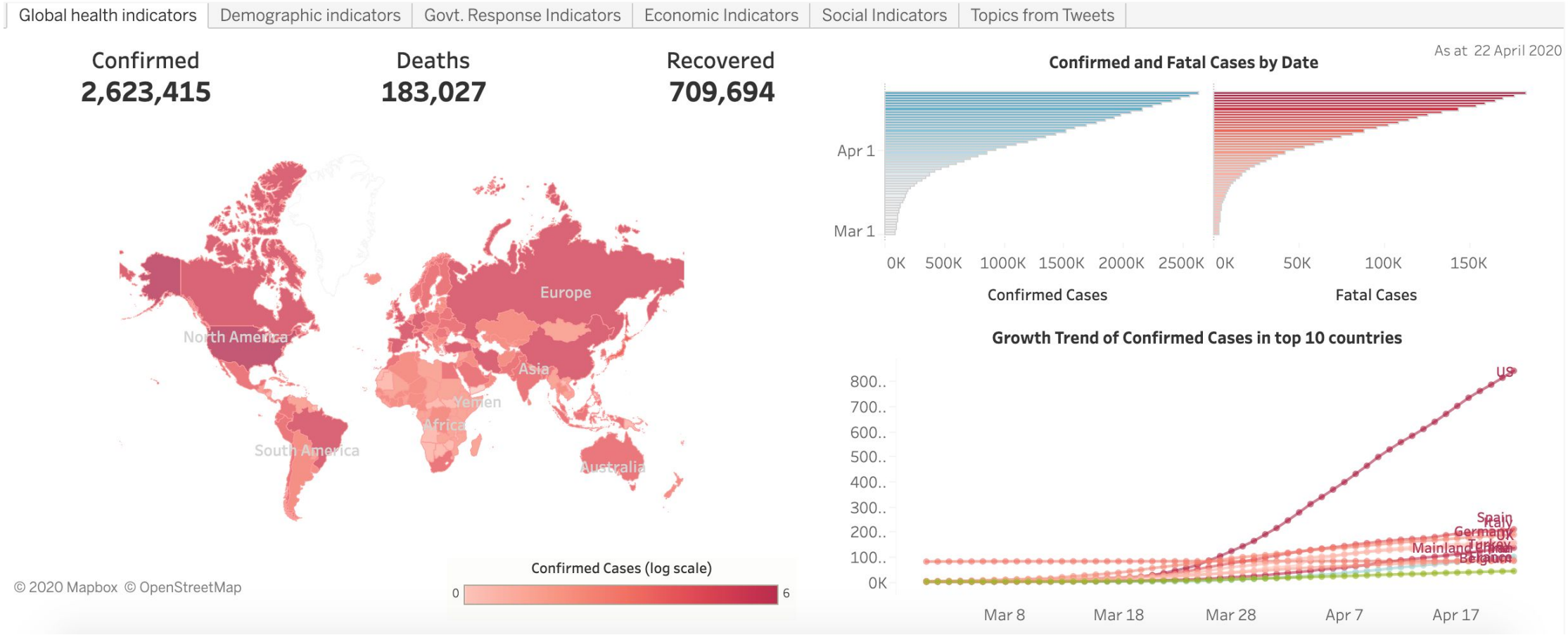
CONFIRMED & FATAL CASES

Forecast for 19-25 April through predictive modeling.

Model

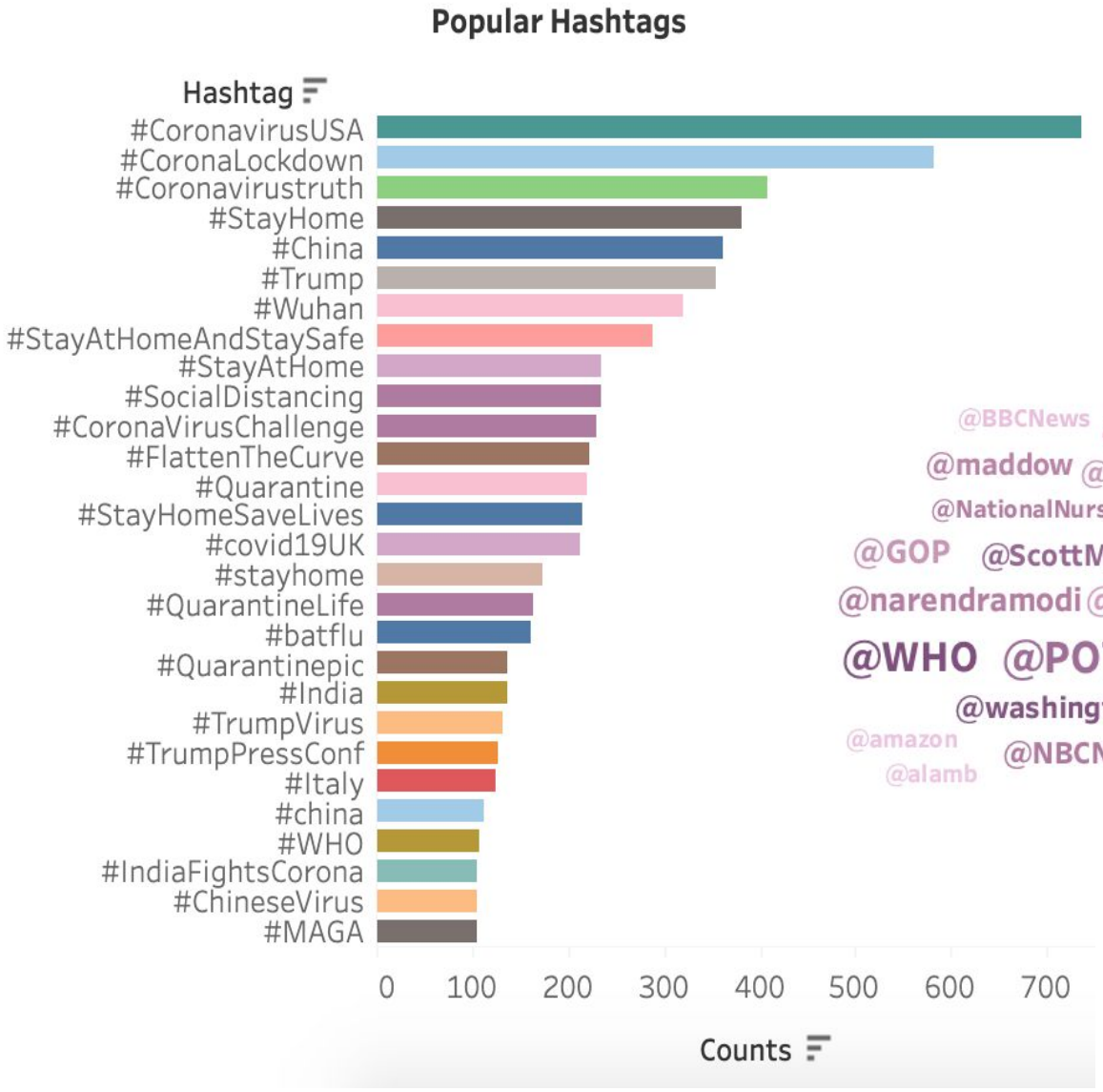
RNN-LSTM

RESEARCH & INSIGHTS



https://public.tableau.com/profile/akhila.joseph#!/vizhome/CovidDashboard2_15860650422580/Globalhealthindicators?publish=yes

SOCIAL PULSE - TWEETS EDA



SOCIAL PULSE - TOPIC MODELING

56,000 tweets

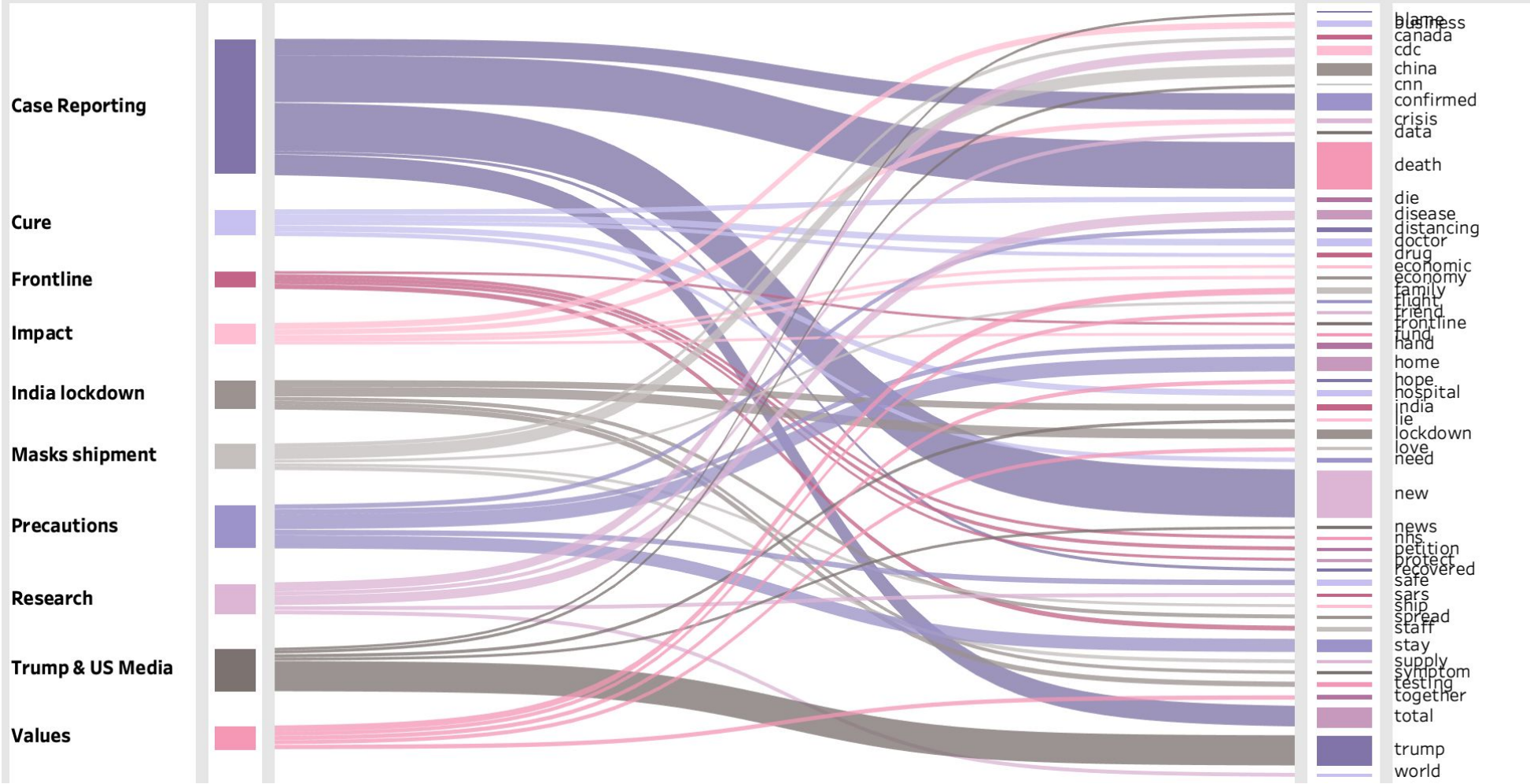
Remove patterns - retweets and mentions → Remove links (http | https → Remove punctuations, numbers & special characters → change to lower case → Tokenize & Lemmatize → Remove stop words → *Normalized corpus*

Normalized corpus → Count - vectorizer → LDA → Evaluate & Visualize

Normalized corpus → TF-IDF → NMF → Evaluate & Visualize

http://localhost:8888/view/Documents/DS/Projects/Akhila-Joseph-Projects/Project_5/Covid-19/code/lda.html

SOCIAL PULSE - TOPIC MODELING



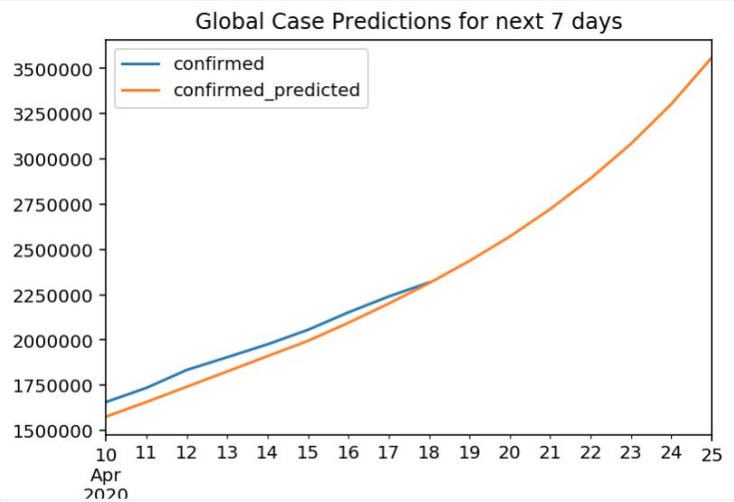
Interpreted Topics

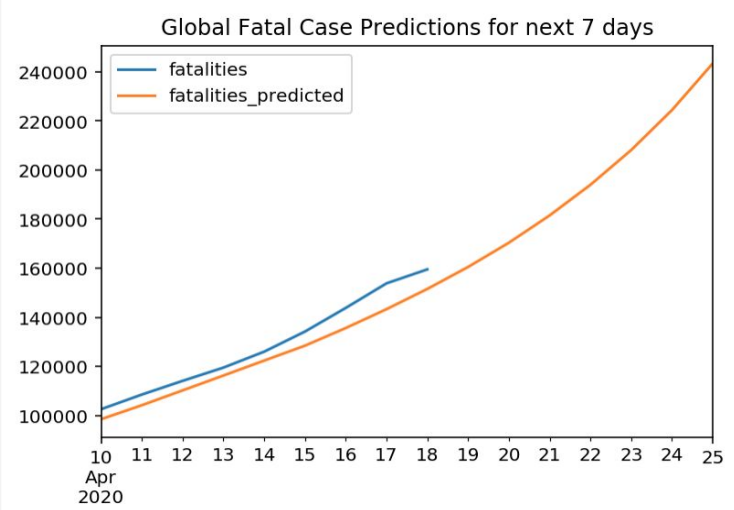
Top weighted words identified by LDA

CASE FORECASTING - MODEL

	confirmed	fatalities
2020-01-22	555	17
2020-01-23	654	18
2020-01-24	941	26
2020-01-25	1434	42
2020-01-26	2118	56

CASE FORECASTING - MODEL PERFORMANCE





	confirmed	confirmed_predicted	fatalities	fatalities_predicted	Avg MAPE: C MAPE	4% F MAPE
4/10/20	1657526	1576827	102525	98456	5%	4%
4/11/20	1735650	1657472	108502	104154	5%	4%
4/12/20	1834721	1741973	114090	110201	5%	3%
4/13/20	1904838	1826314	119481	116253	4%	3%
4/14/20	1976191	1911048	125983	122352	3%	3%
4/15/20	2056054	1995154	134176	128423	3%	4%
4/16/20	2152437	2095004	143800	135609	3%	6%
4/17/20	2240190	2200968	153821	143299	2%	7%
4/18/20	2317758	2314051	159509	151536	0%	5%
4/19/20	2401101	2436595	165043	160489	1%	3%
4/20/20	2472258	2571556	169985	170368	4%	0%
4/21/20	2549293	2723091	176582	181493	7%	3%
4/22/20	2623413	2892294	183025	193969	10%	6%
4/23/20	NaN	3083536	NaN	208135		
4/24/20	NaN	3303240	NaN	224424		
4/25/20	NaN	3559400	NaN	243438		

Confirmed Cases:

MAPE is 3.0%
RMS Log Error 0.04

Fatal Cases:

MAPE is 4.0%
RMS Log Error 0.05

RECOMMENDATIONS & NEXT STEPS

Topic Modeling Enhancements

Next Step

- Use trigram option to enhance the model
- Investigate why bigrams did not return as good a result as desired
- Use grid search to tune hyperparameters

Case Forecasting

Next Step

- Explore country level forecasting
- Possibility to include more features such as weather / temperature conditions, population information etc. for increased accuracy
- Use grid search to tune hyperparameters

DATASETS & REFERENCES

1. *Base dataset: John Hopkins Centre of System Science and Engineering (Covid- case numbers and training data)*
2. *Demographic data for selected countries : The Lancet Digital Health*
3. *Economic Indicators : Yahoo Finance*
4. *Worldwide Government Responses: Oxford University Govt. Response Tracker*
5. *Tweets: Scraped using Twitter scraper API*