KPMG

Machine Learning (ML)-based Prediction Models

For outbreak severity

Sanjaya Sharma, Senior Advisor (Technology), ITA, KPMG

Outbreak Severity Modeling: Summary

We are proposing the use of ML models to predict the severity of the intensity outbreak across the country by state and eventually, by district. To derive the above, we are building stand-alone models that predict the following:

Individual Testing

- Aim is to predict whether a person will be tested positive or negative
- Binary classification (positive/negative)

Positive Cases per State

- Predict number of positive cases on a daily basis using timeseries data, state-wise
- Regression

State Risk Severity Prediction

- Predict if states are at risk or not
- Binary classification

KPMG

© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative, a Swiss entity. All rights reserved.

Document Classification: KPMG Confidential

Model #1: Individual Testing – Approach

Goal: What is the likelihood that a person will test negative or positive?

Data: We've used datasets from the NCDC website:

- Confirmed Cases
- Samples Taken (and tested positive or negative)

To create the following features:

- Gender
- Age: 1-20 / 21-40 / 41-60 / >60
- Risk of the country visited
- Arriving month (arriving India)
- Week of the arriving month
- Source of information
- Delay days
  - Difference between arriving date and observation start date
- State

Approach:
Built a binary classification model, that uses machine learning algorithms like Gradient Boost and Random Forest to identify if each new incoming record is at risk or not

Accuracy:
Precision and recall are measures of accuracy

KPMG

© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative, a Swiss entity. All rights reserved.

Document Classification: KPMG Confidential

Model #1: Individual Testing – Results

Results as of 14th April 2020:

We used a dataset of 34143 cases (7406 positive and 26737 negative), from which we tested the model on 6829 cases.

For positive cases, our model could predict:

- Positive cases with a precision of 91 per cent (from the 1200 cases the model identified as positive, 1095 cases were accurately identified)
- Recall is at about 72 per cent (from 1527 positive cases, we were able to correctly identify 1095 positive cases)

For negative cases, our model could predict:

- Negative cases with a precision of 92 per cent (from the 5629 cases the model identified as negative, 5197 cases were accurately identified)
- Recall is at about 98 per cent (from 5302 negative cases, we were able to correctly identify 5197 positive cases)

KPMG

© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative, a Swiss entity. All rights reserved.

Document Classification: KPMG Confidential

State-level Prediction – Approach

3-day Prediction of Positive Cases (Regression)

- Data: Surveillance data, Samples taken, confirmed cases
  - Attributes such as gender, age, foreign visitor status, number of people under observation, how many are symptomatic and how many are not, and the reporting source
  - Prediction: Predict positive number of cases over a 3-day period
- Algorithms used:
  - Decision Tree
  - Random Forest
  - Gradient Boost (GBM)
- Performance Measures: MAE

Risk Severity Prediction (Classification)

- Data: Date-wise information w.r.t each state (using 6 types of dates present in the surveillance data)
  - Attributes such as gender, age, foreign visitor status, number of people under observation, how many are symptomatic and how many are not, and the reporting source
  - Labels: Using total number of positive cases from internet or surveillance data, a threshold of 3 statistically determined (to identify outbreak class from no-to low outbreak class)
- Algorithms used:
  - SVM Linear Kernel
  - SVM-RBF Kernel
  - Neural Net
  - Logistic Regression
  - **Gaussian Process**
- Performance Measures: Accuracy, F-measure

KPMG

© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative, a Swiss entity. All rights reserved.

Document Classification: KPMG Confidential

Model #2: Positive Cases per State

Prediction over 12th, 13th, 14th April:

To ensure better results over a longer time period, we will need to work with more recent data.

STATE	PREDICTED POSITIVES (12th – 14th April)	ACTUAL POSITIVES (12th – 14th April)
Tamil Nadu	207	235
Rajasthan	219	269
Punjab	46.6	44
Odisha	3	20
Gujarat	67.5	116
Delhi	399	492
Bihar	1.8	3
Uttar Pradesh	113.6	208
Maharastra	251	926

KPMG

© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative, a Swiss entity. All rights reserved.

Document Classification: KPMG Confidential

Model #2: Positive Cases per District (Future)

Forecast over 12th, 13th, 14th April:

We have worked on district-level predictions as well.

However, we will require district-level data, including the number of positive cases confirmed per day, per district for more accurate results.

STATE	DISTRICT	PREDICTED POSITIVES (12th – 14th April)	ACTUAL POSITIVES (12th – 14th April)
Karnataka	Bengaluru	10.27	9
Karnataka	Mysuru	6.09	1
Tamil Nadu	Chennai	25	32
Maharashtra	Mumbai	387	790
Maharashtra	Pune	94	NA

KPMG

© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative, a Swiss entity. All rights reserved.

Document Classification: KPMG Confidential

Model #3: At-risk States – Results

This model trained on data from the 15th of March to the 10th of April. It ranks the states in order of how risk-prone they are for the 11th to 13th of April.

STATE	RISK	SCORE	ACTUAL POSITIVE CASES (11th to 13th April)
Maharashtra	High	1.000	760
Gujarat	High	0.983	231
Madhya Pradesh	High	0.860	169
Andhra Pradesh	High	0.579	64
Uttarakhand	Medium	0.308	0
Haryana	Medium	0.289	8
Chhattisgarh	Low	0.151	21
Jharkhand	Low	0.119	10
Bihar	Low	0.116	5
Assam	Low	0.111	2
Jammu & Kashmir	Low	0.106	63
Himachal Pradesh	Low	0.103	4
Meghalaya	Low	0.090	0
West Bengal	Low	0.086	74
Puducherry	Low	0.072	2
Sikkim	Low	0.053	0
Kerala	Low	0.046	15

KPMG

© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative, a Swiss entity. All rights reserved.

Document Classification: KPMG Confidential

Outbreak Severity Modeling: Focus and Desired Outcome

We are proposing the use of machine learning models to predict the severity of the intensity outbreak across the country by state and eventually, by district.

As we move forward, the outputs of these models can be integrated with existing live dashboards that track the following:

Citizen-facing data

- The current hotspots across the country
- Further details on each hotspot:
  - The number of confirmed cases currently
  - The number of surveilled cases

Administration-facing data

- Emerging hotspots geographically *
- Predicted number of cases that will turn positive *

Facilities data

- Tracking # of facilities, PPEs
- Recommended # of facilities, PPEs *

* Predicted / derived from machine learning-based models

KPMG

© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative, a Swiss entity. All rights reserved.

Document Classification: KPMG Confidential

kpmg.com/socialmedia

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act on such information without appropriate professional advice after a thorough examination of the particular situation.

© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative, a Swiss entity. All rights reserved.

The KPMG name and logo are registered trademarks or trademarks of KPMG International.

Document Classification: KPMG Confidential