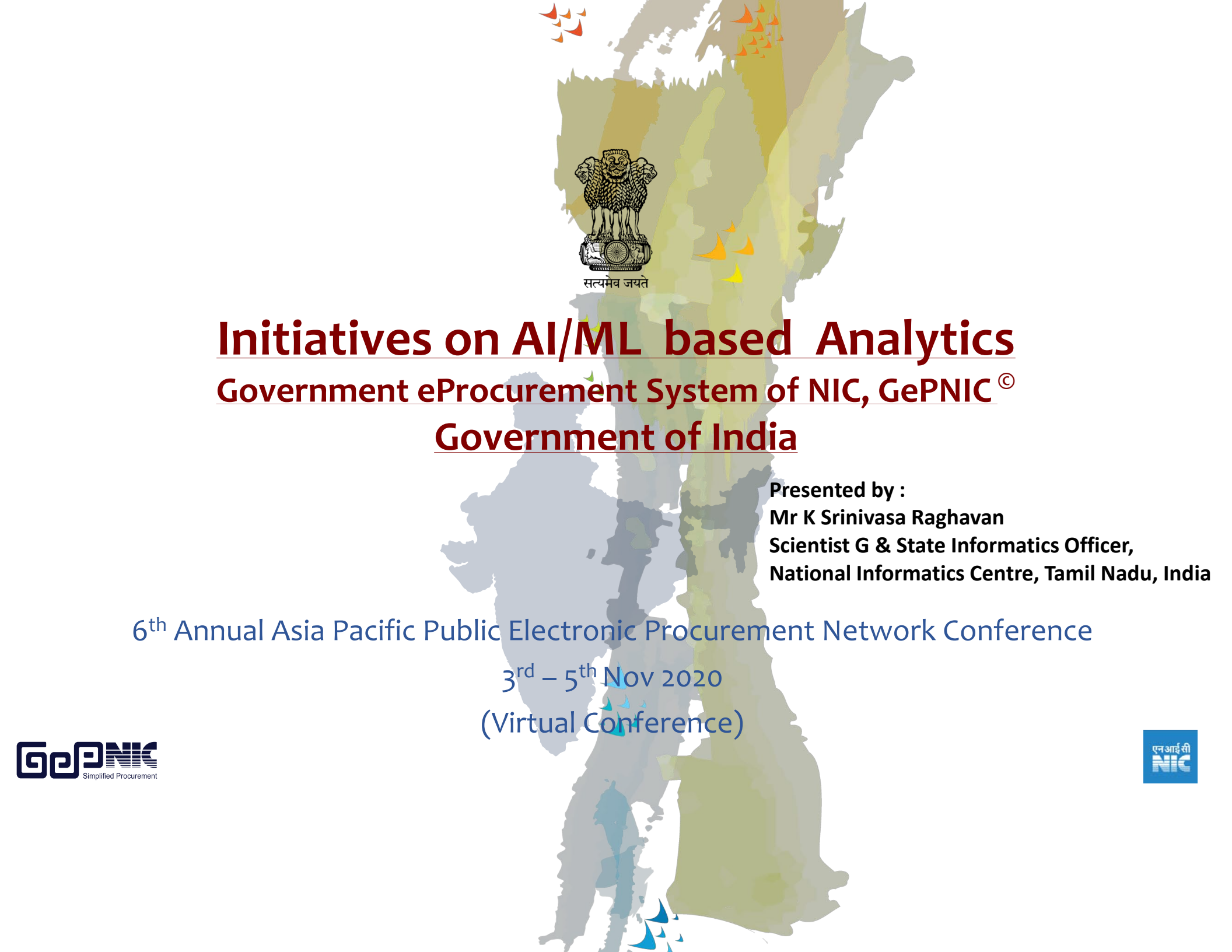


सत्यमेव जयते

Initiatives on AI/ML based Analytics

Government eProcurement System of NIC, GePNIC ©

Government of India

Presented by :
Mr K Srinivasa Raghavan
Scientist G & State Informatics Officer,
National Informatics Centre, Tamil Nadu, India

6th Annual Asia Pacific Public Electronic Procurement Network Conference

3rd – 5th Nov 2020

(Virtual Conference)

GePNIC Simplified Procurement

एन आई सी NIC

Introduction

INDIA

Initiatives to carryout Artificial Intelligence/Machine Learning (AI/ML) based analytics since June 2019.

Few select areas were identified for analysis, as good volume of relevant data is readily available

To start with, a sample set was restricted to 50,000 records for building the model

This model was then applied to data related to tenders published during the period from April 2016 to March 2019 for analysis

Focus Areas for AI/ML analytics

01 **Bidder Segmentation Analysis - based on tender value for which the bidders have participated - to provide focused services to bidders**

Analysis on pattern of association - Bidder Participation trend analysis for pre-emptive Alerts **02**

03 **Forecasting of Volume of Transactions - to envisage and provide Compute and Storage requirements**

Invalid Names - Automatic detection of invalid names entered during user registration process **04**

Focus Areas for AI/ML analytics

01 **Bidder Segmentation Analysis - based on tender value for which the bidders have participated - to provide focused services to bidders**

02 Analysis on pattern of association - Bidder Participation trend analysis for pre-emptive Alerts

03 Forecasting of Volume of Transactions - to envisage and provide Compute and Storage requirements

04 Invalid Names - Automatic detection of invalid names entered during user registration process

Bidder Segmentation Analysis

a) Type of tenders they participate often
b) Products for which they are bidding
c) Category of Tenders
d) Value of tenders

Helps us to segment each bidder to a particular category and this input can be used to for further analysis like send specific Alerts , Predicting participation %

By segmenting each bidder to a particular category, focused services can be planned

Clustering Method is applied to group the Bidders in these categories

K-Prototype Clustering

It belongs to a group of **unsupervised learning** algorithm

Clustering is the task of grouping a set of objects in such a way that objects in the same cluster are more similar to each other than to those in other groups

k-prototypes clustering aims to partition ***n*** **observations** into ***k*** **clusters** in which each observation belongs to the cluster with the nearest centroid.

Elbow Method is used to find the optimal number of clusters.

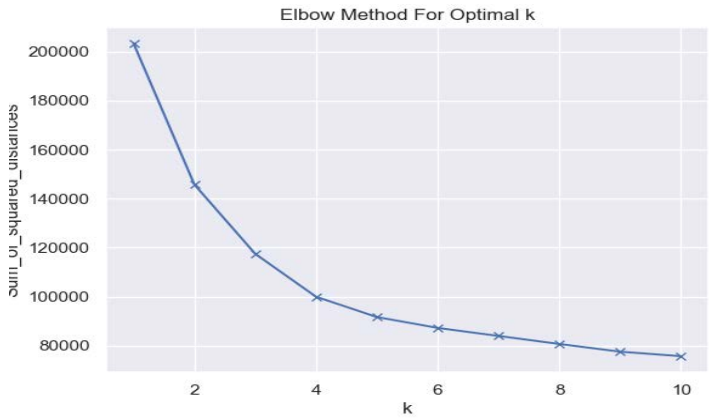
Visualisation of Data Clusters

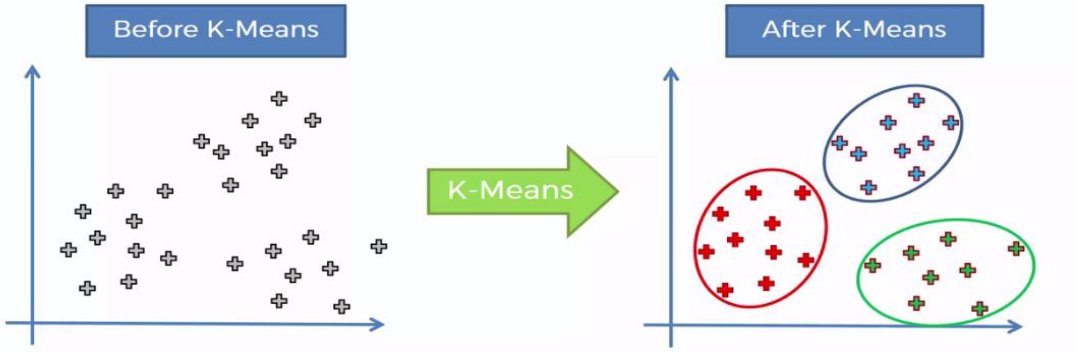
INDIA

An Elbow shape is formed at the 5th cluster point. So, we choose k=5 as our optimal number of clusters.

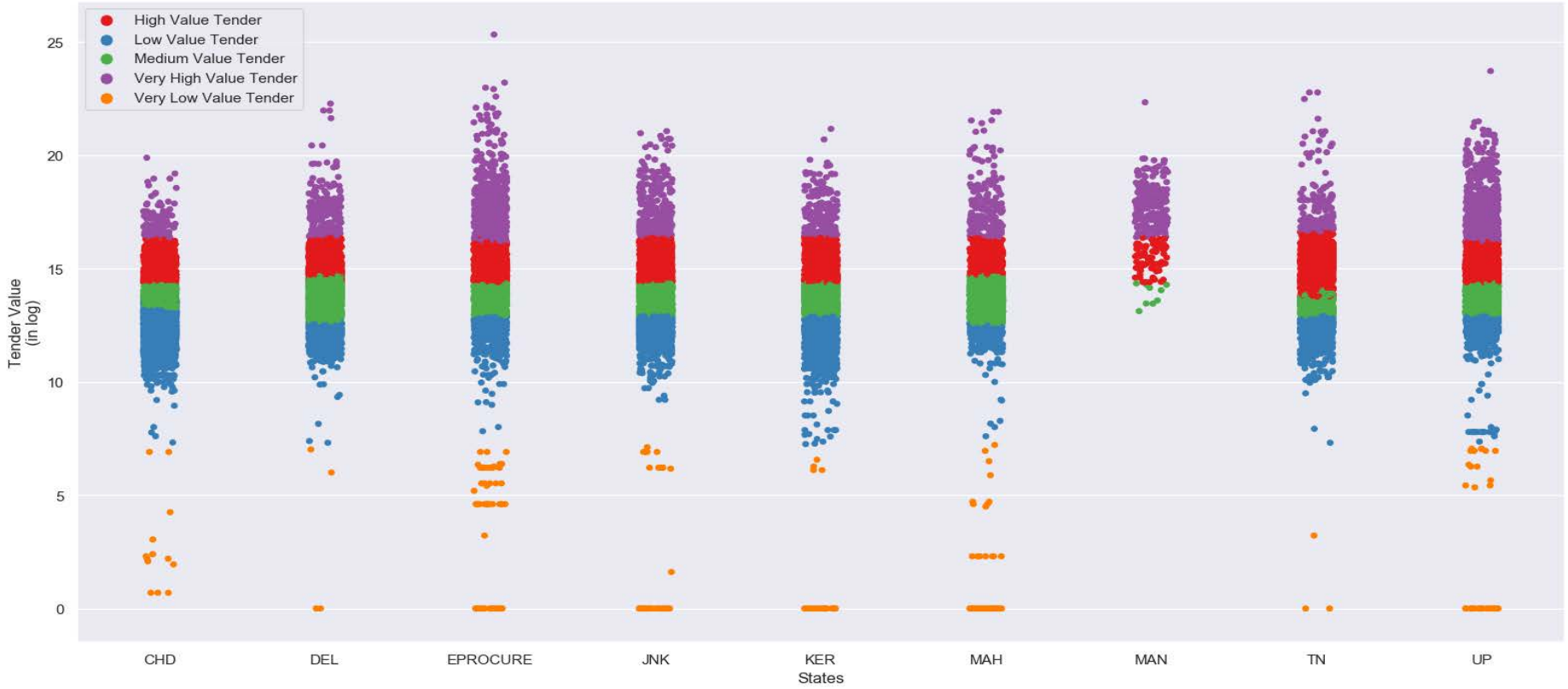
The 5 bidder clusters formed can be interpreted as :

- Very low tender value
- Low tender value
- Medium tender value
- High tender value
- Very high value





Segmentation of Bidders amongst the various provinces/ implementations



Inference drawn from these Segmentation

Inference from the previous graph indicates a clear picture on participation pattern of Bidders in various Provinces - Chandigarh, Delhi, eProcure, J&K, Kerala, Maharashtra, Manipur, Tamil Nadu, UP are given below [applied For Tender Value]

> *High Value Tenders are more in Tamil Nadu State and Medium Value Tenders are less*

> *Very High Value tenders are more in Central Govt & Uttar Pradesh*

> *Very Low value tenders are sporadic in all States*

> *Those bidders who participate in High Value tenders do not participate in low value. (Inference)*

This can be applied to detect the pattern on – Last Hour Bidding , Re-Submissions , Effective Time taken to Bid

Focus Areas for AI/ML analytics

01 Bidder Segmentation Analysis - based on tender value for which the bidders have participated - to provide focused services to bidders

02 Analysis on pattern of association - Bidder Participation trend analysis for pre-emptive Alerts

03 Forecasting of Volume of Transactions - to envisage and provide Compute and Storage requirements

04 Invalid Names - Automatic detection of invalid names entered during user registration process

Analysis on pattern of Association

01 Various platforms like Amazon, Flipkart, Walmart **recommends products** based on the **user behavior**

02 The model is capable of automatically detecting and **recommending tenders** to the bidders based on the **past behavior** the bidders.

03 A **push notification** about such new tender will be sent to the bidders who are highly likely to participate in that.

04 Using this recommendation, the number of bidders participating in a tender can be **increased significantly** as the right bidders are targeted.

05 It also ensures timely automatic alerts to the bidders

APRIORI

INDIA

The *Apriori* Algorithm was proposed by Agrawal and Srikant in 1994.

***Apriori* is an association rule mining algorithm used to identify frequent item sets.**

The three important metrics of apriori are support, confidence and lift.

- **Support -** Percentage of a categories that contains the item or item set.
- **Confidence -** Percentage of times that category B is bidded, given that category A was bidded
- **Lift -** Measure for the relationship between A and B

PRODUCT CATEGORY ASSOCIATION

antecedents	consequents	support	confidence	lift
frozenset({'Computer- Data Processing', 'Miscellaneous Works', 'Computer- H/W'})	frozenset({'Miscellaneous', 'Civil Works - Construction Works', 'Electronics Equipment'})	0.0414707	0.889908	12.2441
frozenset({'Computer- Data Processing', 'Miscellaneous Works', 'Miscellaneous'})	frozenset({'Civil Works - Construction Works', 'Computer- H/W', 'Electronics Equipment'})	0.0414707	0.873874	11.7471
frozenset({'Computer- S/W', 'Miscellaneous Works', 'Civil Works - Construction Works'})	frozenset({'Miscellaneous', 'Computer- H/W', 'Electronics Equipment'})	0.0427533	0.917431	11.7261
frozenset({'Computer- S/W', 'Civil Works - Construction Works', 'Food Products'})	frozenset({'Miscellaneous', 'Computer- H/W', 'Electronics Equipment'})	0.0401881	0.903846	11.5524
frozenset({'Computer- S/W', 'Stationery', 'Civil Works - Construction Works'})	frozenset({'Miscellaneous', 'Computer- H/W', 'Electronics Equipment'})	0.0401881	0.903846	11.5524
frozenset({'Computer- S/W', 'Miscellaneous', 'Food Products'})	frozenset({'Civil Works - Construction Works', 'Computer- H/W', 'Electronics Equipment'})	0.0401881	0.854545	11.4873
frozenset({'Computer- Data Processing', 'Miscellaneous Works', 'Civil Works - Construction Works'})	frozenset({'Miscellaneous', 'Computer- H/W', 'Electronics Equipment'})	0.0414707	0.873874	11.1693
frozenset({'Computer- S/W', 'Food Products'})	frozenset({'Miscellaneous', 'Computer- H/W', 'Electronics Equipment'})	0.0440359	0.872881	11.1567
frozenset({'Computer- S/W', 'Stationery'})	frozenset({'Miscellaneous', 'Computer- H/W', 'Electronics Equipment'})	0.0436084	0.871795	11.1428
frozenset({'Miscellaneous Works', 'Civil Works - Construction Works', 'Food Products'})	frozenset({'Miscellaneous', 'Computer- H/W', 'Electronics Equipment'})	0.0431808	0.87069	11.1287
frozenset({'Computer- Data Processing', 'Miscellaneous Works', 'Miscellaneous', 'Computer- H/W'})	frozenset({'Civil Works - Construction Works', 'Electronics Equipment'})	0.0414707	0.92381	11.081
frozenset({'Medicines', 'Food Products'})	frozenset({'Miscellaneous', 'Computer- H/W', 'Electronics Equipment'})	0.0410432	0.857143	10.9555
frozenset({'Stationery', 'Food Products'})	frozenset({'Miscellaneous', 'Computer- H/W', 'Electronics Equipment'})	0.0410432	0.857143	10.9555

It can be inferred that the higher the 'lift' value, the more the chance of association. i.e. Those bidders who have bid for tenders for 'Computer Data Processing' would also bid for tenders for 'Computer-HW'.

Focus Areas for AI/ML analytics

01 Bidder Segmentation Analysis - based on tender value for which the bidders have participated - to provide focused services to bidders

02 Analysis on pattern of association - Bidder Participation trend analysis for pre-emptive Alerts

03 **Forecasting of Volume of Transactions - to envisage and provide Compute and Storage requirements**

04 Invalid Names - Automatic detection of invalid names entered during user registration process

Forecasting Analytics

SARIMA=**S**easonal + **A**uto**R**egression + **I**ntegration + **M**oving **A**verage

To use SARIMA there are three steps

Define the model

Fit the defined model

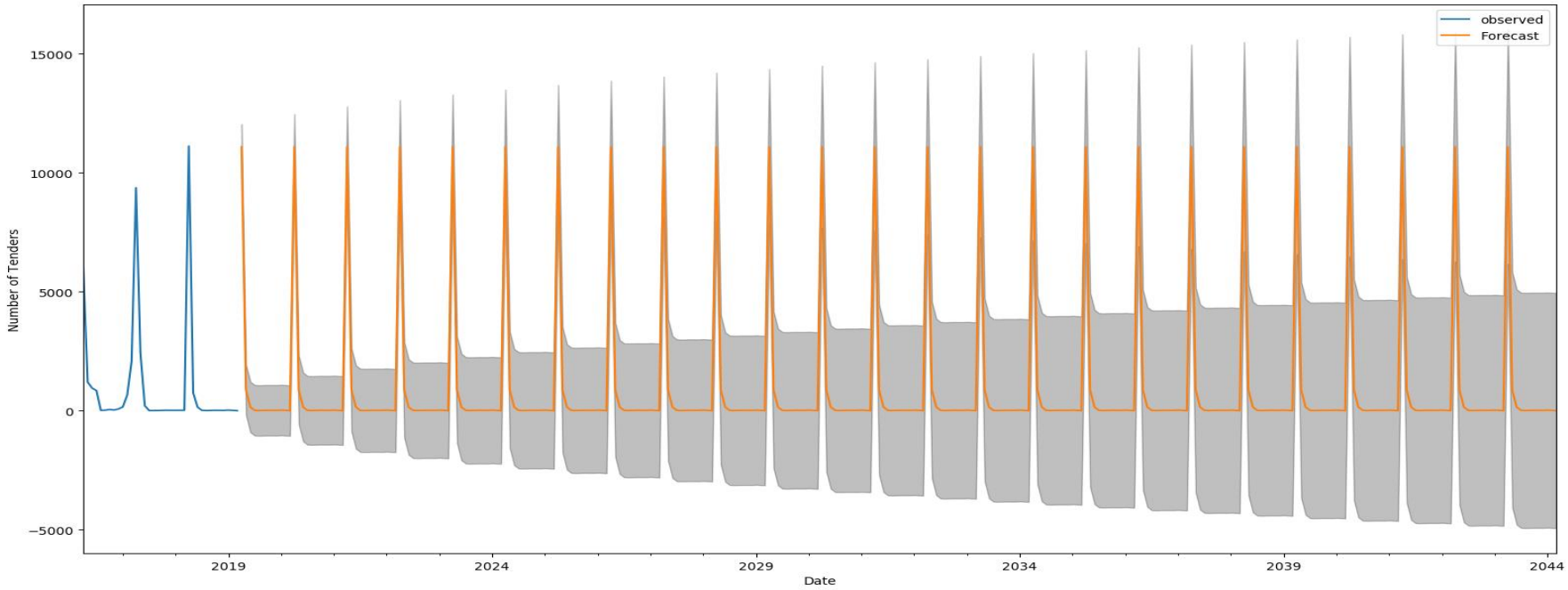
Make a prediction with the fit model

The model can be used to forecast the number of tenders to be handled, Peak time Bids , Peak Month load in future in GePNIC eProcurement platform

A) Forecast using SARIMA model

INDIA

Based on the three years data, the forecast for the next few years have been. It may be noticed that the years in which the actuals are almost in line with the forecasted value can be observed.



Focus Areas for AI/ML analytics

01 Bidder Segmentation Analysis - based on tender value for which the bidders have participated - to provide focused services to bidders

02 Analysis on pattern of association - Bidder Participation trend analysis for pre-emptive Alerts

03 Forecasting of Volume of Transactions - to envisage and provide Compute and Storage requirements

04 **Invalid Names - Automatic detection of invalid names entered during user registration process**

Auto Detection of Invalid Names

01 **To flag the users who have entered invalid names, an intelligent model is created to classify the names.**

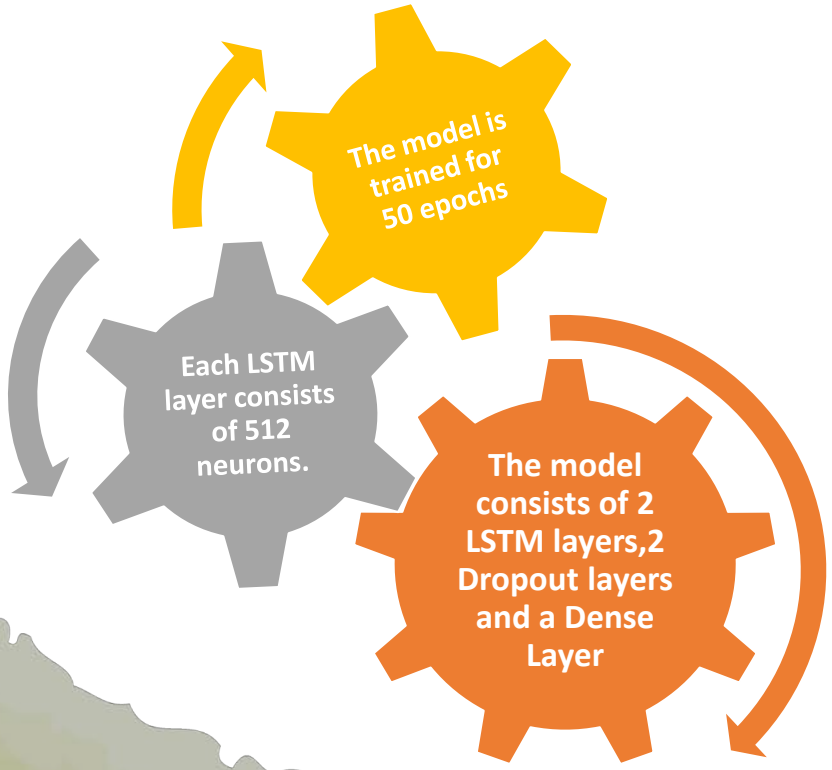
02 **Real names:-Indian and Foreign Names**

03 **Invalid names:-common verbs, adjectives, names generated from a random word generator.**

04 **The size of the total dataset is 2 lakh names**

Long Short-Term Memory
LSTM Model Parameters

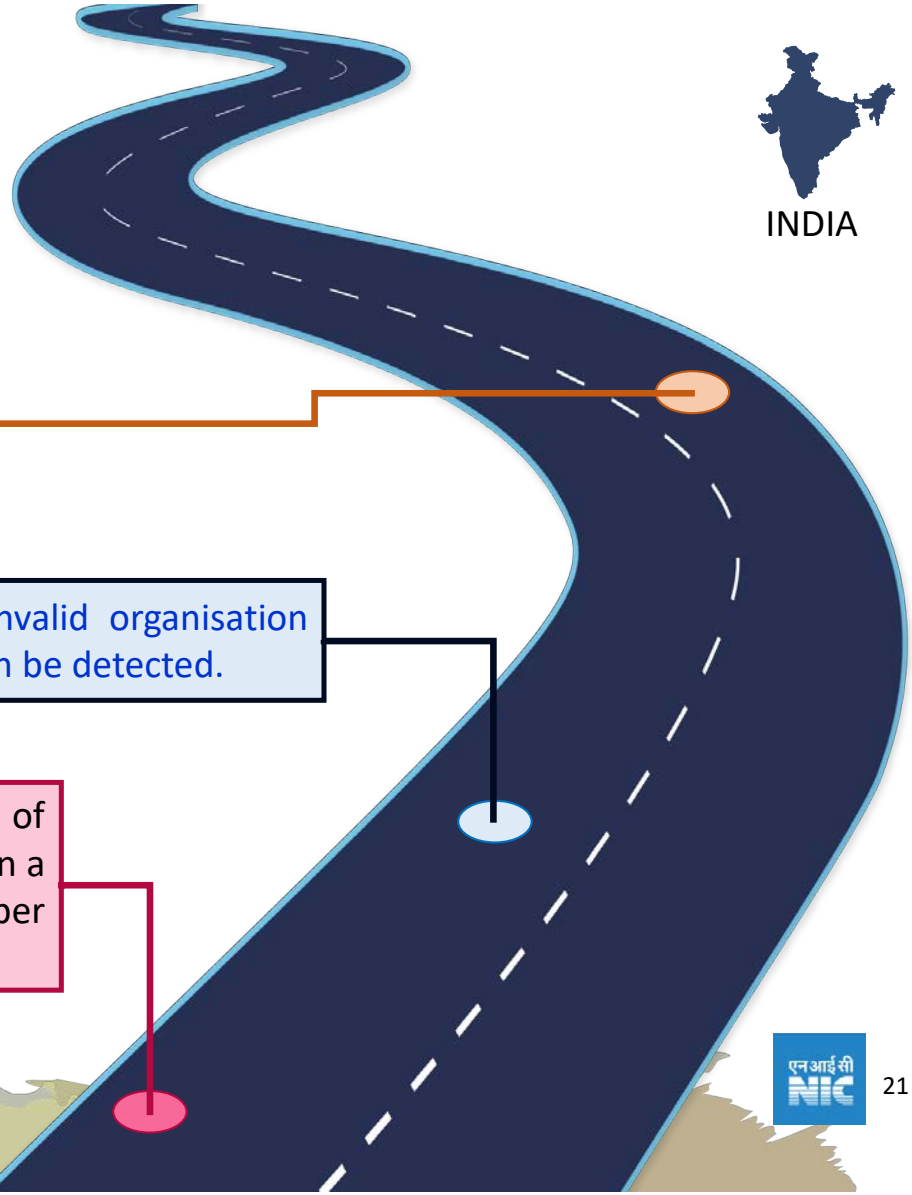
INDIA



Model's Predictions


```
Type the Name
krishna kanth
krishna kanth is a Original Name

Type the Name
rupam
rupam is a Original Name

Type the Name
priya
priya is a Original Name

Type the Name
 gem,m gejkjgnjgn
 gem,m gejkjgnjgn is a Invalid Name

Type the Name
beautiful
beautiful is a Invalid Name

Type the Name
nobody
nobody is a Invalid Name
```

Road Ahead



Various other quantitative measures like number of bidders , number of users, number of corrigendum etc in a particular month can be predicted similar to the number of tender's method.

Similar to the Invalid User Names, Invalid organisation names , product category names etc can be detected.

Various other quantitative measures like number of bidders , number of users, number of corrigendum etc in a particular month can be predicted similar to the number of tender's method.

GePNIC Simplified Procurement

Tools and Libraries used

INDIA

- **Fbprophet-0.5**
- **Keras-gpu-2.2.**
- **Matplotlib-3.1.0**
- **Numpy-1.16.4**
- **Pandas-0.24.2**
- **Scikit-Learn-0.21.2**
- **Seaborn-0.9.0**
- **Statsmodels-0.9.0**

Digital Takeaways :

Please Visit :

- **Central Public Procurement Portal - https://eprocure.gov.in**
- **GePNIC Product Details - https://gepnic.gov.in**
- **Dashboard - https://eprocure.gov.in/eprocdashboard/**

Thank you

Queries may be sent to
ks.raghavan@nic.in

https://www.linkedin.com/in/srinivasa-raghavan-9b557019