

Spatio-Temporal Analysis of Archived Web News for Precise Political Event Detection and Impact Analysis in India's Southern States

Mr. M. Muthamizharasan

AVC College(Autonomous)

R. Ponnusamy (✉ ponnusamyr@citchennai.net)

Chennai Institute of Technology

Research Article

Keywords: Detection Accuracy Rate, Spatio-Temporal Investigation, Net News Archives, Political Events, Financial Stability

Posted Date: April 12th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1516181/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The national political system has the directly impacts the economy. Every political party has their own economic policies. This has the major impact on the economic stability of the nation. Sometimes, few political parties rule either the state or national Governments. Their policies directly reflect in the state or national level financial systems. In general political parties, pay more attention to the state in which they control or aspire to establish their party's rule. Political discourse has an unspoken impact on the state's financial stability. Data from unstructured political events (1200 records) produced for accurate political event prediction; data was pulled from the archives of web news accumulated over the last 5 years to better understand the impact of such impact discovered in this article. This recommended work includes the advantages of Spatiotemporal Analysis of Web News Archives for forecasting network features modeled through political events and converting them into possible information based on natural language processing qualities. The suggested approach constructs a spatial analysis based on hotspots for estimating the sequence of political events in India. Lastly, the resulting data is nurtured into several classifiers for a more reliable prediction and identification procedure. This proposed technique reached a maximum frequency of 78.21 percent and 93.58 percent, respectively, when compared to the Random forest method and the K-neighbor neighbour (KNN) algorithms.

1 Introduction

The political events are considered as the social issue that influences the economy of the country, economic growth of individuals and quality of life [1–2]. It is difficult to comprehend the way of thinking of 'for what reason do things occur?', 'what is the idea of progress?' in the political field. An endeavor at a full study of the pretended by the occasion in way of thinking would nearly be equivalent to a past filled with reasoning itself. In ongoing way of thinking has seen an upsurge of interest in the idea of occasions, in both the logical and the mainland customs. As per the theoretical metaphysics, it has seen the return or reappearance of the occasion to the center stage of philosophical conversation or discourse. The occurrence of political events impacts the economic stability of the country on an international scale [3–4]. The political events occurrence introduces ambiguity to the investors regarding the process of investing huge funds depending on the stability of the government [5]. In order to determine the stability of government with respect to change in political events, algorithms that predict the changes based on location and time (spatial and temporal characteristics) are essential [6]. These algorithms play a vital role in understanding the strategies that need to be implemented during the change of a specific change in a political event [7–9]. In this context, internet-oriented news resources such as new channel archives and online newspapers that have dramatically grown in converges, volume and number possesses potential and authentic data [11–12]. This significant and authentic data provides indispensable information in order to actually visualize the change in the political events happening over a specific geographical area and time period [13–14]. These authentic news archives are not well categorized or arranged, thereby making it highly challenging during the process of determining useful information that are more correlated to interesting change in political events happening in a region [15]. It provided high

valuable information source as it inherits purposeful and rich content as it is carefully captured by specialists, who have more expertise in investigating the change in the political events occurring in a specific region. It also depicts the core dimensions that are highly associated with a particular article [16]. At this juncture, the most authentic and popular archives of India's newspaper such as Indian Express, Times of India, The Hindu, The Hindustan Times, Central chronicle, and. Economic Times can be useful for potential estimation that leads to accurate identification of change in political events [17].

In this paper, Spatio-temporal analysis scheme that concentrates on the objective of identifying the accurate change in political events in India is proposed by utilizing the benefits of news archives that are available free of cost. This spatio-analysis approach extracted key works that depicts the complete representative on behalf of the researchers and news body based on the derivation of Natural language processing. This NLP schemes aids in mining the potential data from the existing news web archives. It identified the high-change political regions by inheriting the benefits of geo-statistical-based approaches. It is proposed for investigating the spatial data represented in diversified dimensions through the incorporation of Geographical Information Systems (GIS) techniques. This GIS-based technique plays an anchor role in exploring and visualizing the possible incidence of political event change through the construction of map layers that correlate to spatial data visualization. This spatial data visualization aided in detecting the trends and patterns of political event changes in India. It is proposed with machine learning and data mining merits that could be possibly applied over the spatial dataset in order to significantly facilitate the determination of change in political events with utmost accuracy. These kinds of spatio-temporal schemes form the major foundation to the companies and individuals in identifying and formulating new policies that could benefits them in the future. It included a information retrieval strategy that extracted important keywords from the news archives and attribute extraction from the headlines of the news through the employment of spatial investigation with added benefits of machine learning to forecast future occurrence of political events in the nation. In specific, Weka tool is used for supporting the process of data mining and employing some specific machine learning algorithms that are essential for achieving this objective. This Weka tool is utilized for achieving the objectives of preprocessing, selection of features, clustering and data classification. This propose schemed utilized a improved kNN algorithm and confirmed better accuracy of 97.21% in predicting the change in political events. This improvement in accuracy during the prediction of political event change is determined to be 8.42%, better than the existing works of the literature considered for investigation. The major contributions of the proposed scheme are listed as follows.

- i) It is proposed as a spatio-temporal analysis approach that significantly predicts the patterns of crimes from the sources of news archival data in order to derive potential information from the news text for determining political progression of BJP in Telangana as an example.
- i) It is propose to support the individuals, companies and various commercial bodies that are in the anticipation of investigating the trends of political event change through the proves of spatial distribution analysis.

- ii) It is proposed with machine learning algorithms for precise prediction of political event change (for example the progression of BJP in different states in India).
- iii) It is also proposed with the added feasibility of employing machine learning and geospatial methods that aids in better prediction of political events using the past five years data available in the archives of the web.

The remaining section of the paper is organized as follows. Section 2 presents the exhaustive review of the existing works of the literature that are highly correlated with the problem considered for study. Section 3 depicts the detailed view of the proposed spatio-temporal analysis scheme propounded for predicting political change in India through the inclusion of machine learning and geo-spatial methods. Section 4 demonstrates the results determined for the proposed spatio-analysis scheme with suitable justifications. Section 5 concludes the paper with major contributions and future scope of research.

2 Literature Survey

Twitter-based spatio-temporal analysis scheme was proposed for high level representation of new events based on social media information [18]. This method of representation included temporal information associated with the location of events' occurrence with respect to geographical entities. This spatio-temporal context-sensitive event representation included the spatial, temporal and social information pertaining to the event representation. It was facilitated with the investigation of world news based on the geo-political and social dimensions. It was proposed with the analysis of international relations and historical event information extraction associated with the news information retrieval tasks. It also used a Galean tool for exploring and retrieving the historical news events that are within the context of temporal and geopolitical contexts. The results of this spatio-temporal analysis scheme confirmed better accuracy in event detection to the maximized level on par with the strategies considered for evaluation. Then, Pollution hot spots-based spatio-temporal analysis scheme was proposed for mining the large degree of air pollution in a specified region based on the merits of time-series analysis and clustering methods [19]. This spatio-temporal analysis was conducted with the air pollution data pertains the ozone and stratosphere layers determined in United Kingdom in the year 2015–2017. The results of the spatio-temporal analysis confirmed better prediction associated with the climatic seasons that could possibly prevail over United Kingdom.

A Modified Auto Regressive Integrated Moving Average (ARIMA)-based crime event analysis scheme was proposed for efficient crime prediction based on the merits of big data technologies [20]. It was proposed to support a different crime trend that generally occurs in different crime locations with some crime selection sites. It was proposed for preventing the limitations of the Generalized Linear Model and Linear model, since these models fails in attaining certainty during crime prediction process. The simulation results of this crime prediction scheme confirmed netter insight into the complexity and scope that could be possibly achieved during crime prediction with the view to improve certainty during the crime event investigation. An Air Quality determination scheme based on spatio-temporal properties was proposed for

ascertaining the factors of air quality in certain regions of China [22]. This air quality estimation scheme was completely based on the datasets of Air Quality Index (AQI) gathered during the years 2014–2016. This spatio-temporal analysis confirmed that the annual mean value of the urban AQI was decreased monotonically from the year 2014. It identified that Northern China was able to retain high AQI value compared to the southern part of China. It included a Moran's I Index for determining the impact of air quality over the adjacent regions. This scheme also included the variations in meteorological and topographical variations that has the possibility of resulting in spatiotemporal variations in the concentration in pollutants.

Further, a spatio-temporal analysis scheme was proposed for determining the trend pattern of rainfall extremes determined in the river basins of Vamsadhara and Nagavali [23]. It employed the Man-kend pollutants all tests that explored the periods comprising post-1950, pre-1950 and long term (1901–2018). It identified the spatial patterns based on the inclusion of kriging interpolation method. It included the method of Sen's slope for determining the magnitude in rainfall and rainfall extremes that are measured in CDD and CWD metrics. This spatio-temporal analysis clearly proves that the decreasing trend is visual with respect to the rainfall extremes both in the pre-and post-1950 period. The accuracy, precision and recall value attained by this approach was improved by 8.21%, 6.72% and 5.48%, better than the baseline shames. Then, a spatio-temporal analysis scheme based on the advantage of geo-statistical methods was proposed for accurate identification of patterns that pertains to the progression of malaria in South Africa [24]. It also included geo-coding spatial analysis for determining the increase in the rise of malaria patients. But, it faced the challenge of handling huge amount of data that need to explore during the spatial-analysis process. The accuracy achieved by this pattern discovery approach is determined to be 93.28%, which is still below par to the benchmarked schemes propounded in the literature. The precision and recall value attained by this approach was improved by 9.36% and 6.59%, better than the benchmarked schemes used for investigation.

Extract of the literature

The existing works of the literature is considered to possess the following shortcomings as listed as follows.

- i) Majority of the proposed schemes failed in utilizing the actual benefits of geo-spatial coding schemes for better prediction of patterns necessitated in a specific problem domain.
- ii) Most of the works contributed to the literature were not capable in potential extraction of significant information, since they faced the challenge of big data investigation during the exploration of news web archives.
- iii) The accuracy, precision and recall facilitated by the available works of the literature still possessed a possible room of improvement.

iv) The extraction and investigation complexity inherent with the existing works of the literature also needs significant and phenomenal improvement.

3. Proposed Precise Political Event Detection Using Spatio-temporal Analysis Scheme

This proposed Precise Political Event Detection using Spatio-Temporal Analysis (PPED-STA) Schemes proposed by political event mapping facilitated through the extraction of data from the archives of web news. In the complete extraction process, the news records determined from the archives are processed through the support of Python processing modules that increases the possibility of deriving valuable information from the text of news records. In addition, the spatial characteristics of political events are dispersed, random or clustered in order to explore the influence of political events over the nation financial stability.

3.1 Data collection and Processing

The complete data pertaining to the proposed scheme are comprehensively achieved by crawling the data through the data miner tool over the archives of web news of most of the well known Indian newspapers such as Indian Express, Times of India, The Hindu, The Hindustan Times, Central chronicle, and. Economic Times. In specific, the data miner tool is used for complete data crawling processes as it is capable of collecting data from a particular website and transforms into a tabular form. This tool is also potent is categorizing the news archive based on the URL, date, description, title, etc. The news records' attributes considered in the proposed scheme are location, latitude, longitude, URL, title and description. The news data archives considered for the study ranges between 2016–2019, which are further mined for the formulation of the research model. In total, 1120 records associated with the political events of Ayodha Issue were extracted for further process of screening that concentrated on removal of duplicates. Out of 1120 records, it is identified that 20 records were duplicated as they are extracted multiple times and, hence they are removed. The data comprises of 1100 records and the total number of records pertaining to each category is depicted in Table 1.

Table 1
Different Category of Political Events Records, Precision and Recall

Type of political event	Number of records	Precision	Recall
BJP Success in Municipality elections in Telangana	350	81.16%	72.18%
BJP wins 41 out of 59 by-poll seats in Gujarat	210	84.84%	75.82%
BJP retains six seats in Uttar Pradesh	150	82.39%	74.81%
BJP has its first-ever electoral success in Kashmir	100	83.28%	75.68%
BJP and Nitish win the Bihar elections	80	88.19%	78.56%
BJP set win in Manipur	70	87.26%	81.26%
BJP secures more seats in Madhya Pradesh	50	89.32%	84.28%
BJP bags maximized seats in Karnataka	40	87.48%	87.24%
BJP takes over one seat from Telangana Rashtra Samithi	30	85.72%	86.41%
BJP gains its new rise in West Bengal	20	85.68%	84.59%

In Table 1, the number of records associated with Ayodha Issue (450 records) are highest among all. In specific, BJP IIS regaining the parliament comprises of 350 records, Demonitization Issue of 150 records, Farmer Agricultural Policy of 125 records, BJP election agenda of 80 records and of Regional intolerance Policy of 70 records, respectively. Moreover, the data are considered to be further processed based on the diversified modules of Python that includes panda, geopy, natural language toolkit and regular expression for normalizing the data. Further, the natural language toolkit is capable of performing the process of topic segmentation, sentimental analysis, tokenizing, stemming and tagging of part-of-speech named entity recognition. Further, it is significant in exploring, preprocessing and perceiving the written text. It also aids the computer in significant text interpretation process. In addition, the natural language toolkit is used in this study as it incorporates named entity recognition process that defined the classes of information extraction. The, regular expression module facilitates superior accuracy and it is considered to be better than the N-gram methodology even though they are considered for information extraction from the text in most of the existing studies. The modules of regular expression are considered for determining the political event prediction are, i) BJP Success in Municipality elections in Telangana, ii) BJP wins 41 out of 59 by-poll seats in Gujarat, iii) BJP retains six seats in Uttar Pradesh, iv) BJP has its first-ever electoral success in Kashmir, v) BJP and Nitish win the Bihar elections, vi) BJP set win in Manipur, vii) BJP secures more seats in Madhya Pradesh, viii) BJP bags maximized seats in Karnataka, ix) BJP takes over one seat from Telangana Rashtra Samithi and x) BJP gains its new rise in West Bengal. Thus, the regular expression modules are used as it is potent in deriving the best results through parsing text.

3.2 The process of Geo-coding

In the process of spatial data analysis, the geo-coding is considered to facilitate a indispensable role and assigns the associated longitude and attitude to the actual location for the purpose of better prediction and visualization of political patterns. This process of geo-coding is achieved using geo-py existing in Python. In this proposed scheme implementation, the geopy module 2.0.0 together with the Python Pandas module is used geo-coding, since it is has the potentiality in handling the rows and columns during the geo-coding process. The complete process involved in the process of entire information extraction is presented in Fig. 1.

3.3 Removal of Punctuations

Most of the Python modules are not capable of text processing, if it contains the punctuation. Taking this into account, the punctuation present in the title and description pertaining o the news records at the first step are removed in order to facilitate the algorithm to work well over it.

3.4 Information Extraction Evaluation

In this proposed scheme implementation, the parameters of precision and recall are utilized for evaluating the accuracy of the political information extraction, particularly the occurrence of political event as presented in Table 1. Moreover, the attributes and characteristics of the data are presented in Table 2, which highlights the Location, Latitude, Longitude, Type of political event, Political event pattern recognized, Event date and Title pertaining to the political event occurrence.

Table 2
Data Attributes

Data	Type of data	Description	Examples
Location	String	Location of the political event occurrence	Telangana
Latitude	Float	Latitude	45.632
Longitude	Float	Longitude	56.921
Type of political event	String	The political event occurrence type	BJP Progression
Political event pattern recognized	Integer	Number of seats won by BJP	5 Seats won by BJP
Event date	Date	Date of the event occurred	21, December, 2020
Title	String	Title of the political event	Five Seats in Telangana is won by BJP

3.5 Preprocessing

In the preprocessing step, the complete attributes are extracted and the comprehensive set of all instances are integrated for the available description, Then, the process of data cleaning is achieved with

the description and URL attributes removed as the prediction can be also facilitate without these fields. In this case, the attributes considered for preprocessing is completely determined based on extensive literature exploration. This step of preprocessing did not included the requirement of minimizing the number of parameters, hence none of the data reduction techniques were used during this step. The data attributes that are preprocessed are considered as the acceptable format for being fed into the Naive Bayes, KNN and random forest algorithms. Hence, additional data transformation and steps of discretization are not applied keeping in mind of the acceptable degree of the preprocessed data. However, only one potential encoding scheme is employed for converting the preprocessed data which is in categorical attributes into its numeric format.

3.6 Geo-Spatial Data Mapping

In this step of Geo-Spatial Data Mapping, data visualization which is considered as the visual art is utilized for representing the information and data in a more graphical manner. This Geo-Spatial Data Mapping-based data visualization scheme is employed for exploring the patterns and trends of data in a more peculiar manner. In specific, this Geo-Spatial Data Mapping, data visualization scheme is used in the proposed scheme for exploring diversified number of political events with the aid of the map that plays an anchor role in predicting the futuristic political events.

Moreover, the modules and tools of ArcGIS are utilized for facilitating information display on the map.

3.7 Arcgis-based political event data visualization

The tools and modules of ArcGIS used for political event data visualization is well known for the objective of visualizing spatial datasets. In this proposed scheme, the shapefile of political event dataset is loaded together with the dataset extracted by the ArcGIS in order to represent political event records based on the longitude and latitude. The aforementioned Fig. 2 depicts the ArcGIS representation of the dataset and indicates that BJP is emerging in the state of Telangana from the recent two years. In addition, ArcMap software is used for exhibiting the political event distribution based on different types used for prediction. In addition, ArcMap software facilitates different categories of geographical representation that clearly demonstrates the regions or areas that are highly susceptible to a specific political event change in the near future can be easily identified.

4. Results And Discussion

The performance of the proposed PPED-STA Scheme is evaluated based on the evaluation metrics of accuracy, precision, recall and F-measure. The accuracy metrics is defined as the potential in predicting labels associated with categorical class. In other words, It is computed as the proportion of correctly predicted instances as specified in Eq. (1)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision refers to the closeness measures of instances with one another and it is computed based on Eq. (2)

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall is defined as the exact positive instances in the dataset that have been accurately determined as positive by the utilized classifier as calculated based on Eq. (3)

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

In addition, the F-Measure is computed based on the weighted harmonic mean of the recall and precision as mentioned in Eq. (4)

$$F - \text{Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Where, TP-True Positive-If the instance is positive and the outcome of the classification is also positive

TN-True Negative-If the instance is negative and the outcome of the classification is also negative.

FN-False Negative-If the record is positive and the outcome of the classification is also negative.

FP-False Positive-If the data record is negative and the outcome of the classification is also positive.

The political event prediction is considered as one of the most significant tasks, particularly when the data availability with respect to political events are not available up to the marks. At the juncture, electronic media is determined as the most powerful tool that can facilitate accurate data and retains to be potential during the research process. Further, data mining tools helped in handling the data and transform into an understandable format which helps in extracting information in order to answer political event patterns and its relationships. In this prediction scheme, three machine learning algorithm such as kNN, random forest and Markov Property-based Random Forest were used for the political event prediction process based on archive datasets. The results of the aforementioned machine learning algorithms are investigated and compared with respect to the evaluation metrics of accuracy and prediction.

Table 3 and Fig. 3 presents the accuracy and precision of the proposed KNN-based PPED-STA scheme with different value of k (k set to 3, 5, 7 and 9). The accuracy of the proposed KNN-based PPED-STA scheme with k = 9 performed better than at the remaining values considered for investigation, since the process of discrimination imposed for the collected data played an anchor role in improving accuracy. In contrast, the precision of the proposed PPED-STA scheme with k = 3 performed better as it incorporated better determination of relevant and irrelevant classes from the dataset considered for exploration. The value of the accuracy facilitated by the proposed KNN-based PPED-STA scheme at k = 9 is also maximum upto 95.4%, which is comparatively increased at a mean rate of 4.86% on par with the other values

considered for investigation. Moreover, value of the precision attained by the proposed KNN-based PPED-STA scheme at $k = 3$ is maximized with a value of 94.2%, which is comparatively increased at a mean rate of 5.62% on par with the other values considered for investigation.

Table 3
Results of the proposed PPED-STA:
Accuracy and Precision based on k-NN
attained using Weka tool

Value of k	Accuracy	Precision
3-NN	0.912	0.942
5-NN	0.948	0.936
7-NN	0.953	0.924
9- NN	0.954	0.918

Table 4 and Fig. 4 demonstrates the recall and F-measure (macro and micro-averaged) value of the KNN-based PPED-STA scheme with different value of k (k set to 3, 5, 7 and 9). The recall of the proposed KNN-based PPED-STA scheme with $k = 3$ is identified to provide excellent performance on par with the remaining values considered for investigation. This potential performance is mainly due to the determination of relevant data from the complete set of data considered for exploration. The recall value of the proposed KNN-based PPED-STA scheme at $k = 9$ is proved to be maximum upto 91.2%, which is comparatively improved by 8.32% over the other values considered for investigation. The F-measure (macro-averaged) attained by the proposed PPED-STA scheme with $k = 9$ was identified to be significant as the features considered for computing mean precision and recall is reliable in properly estimating the regression degree available between the data. Thus, F-measure (macro-averaged) attained by the proposed KNN-based PPED-STA scheme at $k = 9$ is also maximum upto 91.2%, which is comparatively improved by 7.24%, better than the other values considered for investigation. The F-measure (micro-averaged) achieved by the proposed PPED-STA scheme with $k = 3$ is confirmed to be predominant as the temporal classification involved during analysis aided in better exploration of the data. Hence, F-measure (micro-averaged) achieved by the proposed KNN-based PPED-STA scheme at $k = 3$ is maximized with 91.8%, which is enhanced by 6.92%, superior to the other baseline values considered for investigation.

Table 4
Results of the proposed PPED-STA: Accuracy and Precision based on k-NN attained using Weka tool

Value of k	Recall	F-measure	
		Macro-Averaged	Micro-Averaged
3-NN	0.912	0.842	0.918
5-NN	0.908	0.854	0.904
7-NN	0.903	0.886	0.896
9- NN	0.896	0.912	0.893

Table 5 and Fig. 5 depicts the accuracy and precision of the proposed k-Random Forest-based PPED-STA scheme with different value of k (k set to 10, 20, 30 and 40). The accuracy and precision of the proposed k-Random Forest-based PPED-STA scheme with k = 10 performed better, since it aided in predominant exploration of spatial and temporal features existing in the dataset considered for study. The value of the accuracy facilitated by the proposed k-Random Forest-based PPED-STA scheme at k = 10 maximized with a value of 68.2%, which is comparatively increased at a mean rate of 6.84% on par with the other values considered for investigation. Moreover, value of the precision attained by the proposed k-Random Forest-based PPED-STA scheme at k = 10 is maximized with a value of 62.2%, which is comparatively increased at a mean rate of 4.59% on par with the other values considered for investigation.

Table 5
Results of the proposed PPED-STA: Accuracy and Precision based on Random Forest attained using Weka tool

Value of k	Accuracy	Precision
10-Forest	0.682	0.622
20-Forest	0.676	0.618
30-Forest	0.672	0.604
40-Forest	0.668	0.602

Table 6 and Fig. 6 presents the recall and F-measure (macro and micro-averaged) value of the k-Random Forest-based PPED-STA scheme with different value of k (k set to 10, 20, 30 and 40). The recall of the proposed k-Random Forest-based PPED-STA scheme with k = 10 is excellent as it adopted integrated feature selection and classification process. The recall value of the proposed k-Random Forest-based

PPED-STA scheme at $k = 10$ is enured is proved to be maximum upto 71.2%, which is comparatively improved by 5.68% better than the other values considered for investigation. The F-measure (macro-averaged) attained by the proposed k -Random Forest-based PPED-STA scheme with $k = 10$ is proved to be potential as the comprehensive set of features considered for prediction is highly optimized before classification. F-measure (macro-averaged) attained by the proposed k -Random Forest-based PPED-STA scheme at $k = 10$ is also maximum upto 56.2%, which is comparatively improved by 6.42%, better than the other values considered for investigation. The F-measure (micro-averaged) achieved by the proposed k -Random Forest-based PPED-STA scheme with $k = 10$ is confiemed to be predominant as the process of spatial and temporal classification involved during analysis aided is completely context-based and it is also capable in handling missing data. Hence, F-measure (micro-averaged) achieved by the k -Random Forest-based PPED-STA scheme at $k = 10$ is maximized with 61.2%, which is enhanced by 5.94%, superior to the other baseline values considered for investigation.

Table 6
Results of the proposed PPED-STA: Accuracy and Precision based on Random Forest attained using Weka tool

Value of k	Recall	F-measure	
		Macro-Averaged	Micro-Averaged
10-Forest	0.712	0.562	0.612
20-Forest	0.706	0.541	0.604
30-Forest	0.696	0.536	0.596
40-Forest	0.684	0.526	0.582

Table 7 and Fig. 7 presents the accuracy and precision of the proposed PPED-STA scheme under the integration of k NN and k -Random Forest classifier. The accuracy and precision of the proposed PPED-STA scheme with 3-NN and 10-Forest is superior, since it combined the possible benefits of k NN and random forest to the expected level. The value of the accuracy facilitated by the proposed PPED-STA scheme with 3-NN and 10-Forest achieved a value of 73.2%, which is increased at a mean rate of 5.12%, better than the other k -NN and k -foresr values considered for investigation. Moreover, value of the precision attained by the proposed PPED-STA scheme with 3-NN and 10-Forest is maximized with a value of 78.3%, which is comparatively increased at a mean rate of 6.42%, better than the other k -NN and k -foresr values considered for investigation

Table 7
Results of the proposed PPED-STA: Accuracy and Precision based on Random Forest attained using Weka tool

Value of k and n	Accuracy	Precision
3-NN and 10-Forest	0.732	0.783
5-NN and 20-Forest	0.729	0.772
7-NN and 30-Forest	0.718	0.764
9 NN and 40-Forest	0.711	0.721

Table 8
Results of the proposed PPED-STA: Accuracy and Precision based on kNN and k-Random Forest attained using Weka tool

Value of k	Recall	F-measure	
		Macro-Averaged	Micro-Averaged
3-NN and 10-Forest	0.852	0.671	0.664
5-NN and 20-Forest	0.841	0.664	0.652
7-NN and 30-Forest	0.834	0.661	0.641
9 NN and 40-Forest	0.826	0.652	0.638

Table 8 and Fig. 8 presents the recall and F-measure (macro and micro-averaged) value of the proposed PPED-STA scheme under the integration of kNN and k-Random Forest classifier. The recall of the proposed PPED-STA scheme with 3-NN and 10-Forest is determined to facilitate remarkable performance, since the merits of k-NN and random forest are proportionally combine for retrieving huge amount of features that aids in better exploration. The recall value of the proposed PPED-STA scheme with 3-NN and 10-Forest is identified to be maximum with a value of 85.2%, which is comparatively improved by 6.56%, superior than the other values considered for investigation. The F-measure (macro-averaged) attained by the proposed PPED-STA scheme with 3-NN and 10-Forest is proved as significant because of the reduction in the false positive rate determined during discriminate analysis. Thus, F-measure (macro-averaged) attained by the proposed PPED-STA scheme with 3-NN and 10-Forest has a maximum value of 67.1%, which is comparatively improved by 5.92%, better than the other values considered for investigation. The F-measure (micro-averaged) achieved by the proposed PPED-STA scheme with 3-NN and 10-Forest is proved to be highly important as it explored diversified dimensions of data with different aspects of exploration. Hence, F-measure (micro-averaged) achieved by the proposed PPED-STA scheme with 3-NN and 10-Forest is maximized up to 66.4%, which is enhanced by 5.86%, predominant over the other benchmarked values considered for investigation.

5. Conclusion

This proposed PPED-STA Scheme is a reliable spatio-temporal analysis approach that significantly predicts the patterns of events from the sources of news archival data in order to derive potential information from the news text for determining political progression of BJP in Telangana. It was proposed for helping the individuals, companies and various commercial bodies that are in the anticipation of investigating the trends of political event change through the proves of spatial distribution analysis. It was proposed with machine learning algorithms for precise prediction of political event change (for example the progreesion of BJP in different states in India). It also included the added advantage of feasibility of employing machine learning and geospatial methods that aids in better prediction of political events using the past five years data available in the archives of the web. The experimental results of the proposed PPED-STA Scheme, on an average improved accuracy by 8.21%, precision by 12.68%, recall by 10.29%, better than the benchmarked approaches considered for investigation. As a plan of future scope, it is planned to formulate a deep learning approach using Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM)-based political event prediction scheme and compare with the currently proposed scheme of this research and determine the similarities and difference in potentialities between the two spatio-temporal analysis scheme.

Declarations

Author Contributions:

Conceptualization: M. M. & R. P.; Methodology: M. M. & R. P.; Validation: M. M. & R. P.; Formal Analysis: M. M. & R. P.; Investigation: M. M. & R. P.; Resources: M. M. & R. P.; Data Curation: M. M. & R. P.; Writing original draft preparation: M. M. & R. P.; Writing review and editing: M. M. & R. P.; Visualization: M. M. & R. P.; Supervision: R. P. Project administration: M. M. & R. P.; Funding acquisition: M. M. & R. P. All authors have read and agreed to the published version of the manuscript.

Acknowledgement: We deeply acknowledge Chennai Institute of Technology for supporting this study through Chennai Institute of Technology Researchers Supporting Project Number (CIT/CAIR/2022/RP-015), Chennai Institute of Technology, Chennai, India.

Funding Statement: This research is funded by Chennai Institute of Technology, CIT/CAIR/2022/RP-015.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Diego Bueso, Maria Piles and Gustau Campus Valls, "Nonlinear complex PCA for spatio-temporal analysis of global soil moisture", *IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, pp. 5780–5783, 2018.

2. Bueso, D., Piles, M., and Camps-Valls, G., Nonlinear PCA for spatio-temporal analysis of earth observation data. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 8, 5752–5763, 2020.
3. Rejichi, S., and Chaabane, F., "Spatio-temporal regions' similarity framework for VHR satellite image time series analysis", 2016 *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2845–2848, 2016.
4. Rejichi, S., Chaabane, F., and Tupin, F, Expert knowledge-based method for satellite image time series analysis and interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 8, No. 5, 2138–2150, 2015.
5. Shoaib Azmat, Linda Wills and Scott Wills, "*Spatio-temporal multimodal mean*", 2014 Southwest Symposium on Image Analysis and Interpretation, San Diego, CA, pp. 81–84, 2014.
6. O. Dossel, T. Oesterlein, L. Unger, A. Loewe, C. Schmitt, et. al., "Spatio-temporal Analysis of Multichannel Atrial Electrograms Based on a Concept of Active Areas", 2018 *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, pp. 490–493, 2018.
7. V. Barone, E. Maranesi and S. Fioretti, "Integration of smartphones and webcam for the measure of spatio-temporal gait parameters", 2014 *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL, pp. 5948–5951, 2014.
8. Bin Liu, Wujiao Dai, Wei Peng and Xiaolin Meng, "Spatio-temporal analysis of the land subsidence in the UK using Independent Component Analysis", 2014 *Third International Workshop on Earth Observation and Remote Sensing Applications (EORSA)*, Changsha, pp. 294–298, 2014.
9. W. Chu and Y. Guan, "Spatio-Temporal motion analysis based suspicious behavior detection", 2016 *13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, Chengdu, pp. 179–183, 2016.
10. H. Hayashi, A. Asahara, N. Sugaya, Y. Ogawa and H. Tomita, "Spatio-temporal similarity search method for disaster estimation", 2015 *IEEE International Conference on Big Data (Big Data)*, Santa Clara, CA, pp. 2462–2469, 2015.
11. A. John, M. Sugumaran and R. S. Rajesh, "Performance analysis of the past, present and future indexing methods for spatio-temporal data", 2017 *2nd International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, pp. 645–649, 2017.
12. Toppireddy, H. K., Saini, B., and Hada, P. S., "Academic enhancement system using EDM approach", 2019 *Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, Vol. 2, No. 1, 56–68. 2019.
13. Babakura, A., Sulaiman, M. N., and Yusuf, M. A. "Improved method of classification algorithms for crime prediction", 2014 *International Symposium on Biometrics and Security Technologies (ISBAST)*, Vol. 2, No. 1, 2014.
14. Xue, Y., and Brown, D. E., "Spatial analysis with preference specification of latent decision makers for criminal event prediction", *Decision Support Systems*, Vol. 41, No.3, 560–573, 2006.

15. Xi Zhou, F. Wang, Chaolin Wang, Xiaocui Zheng, Fuda Zheng, et al., "A novel spatio-temporal similarity algorithm adapted to typhoon disaster cases", *2015 23rd International Conference on Geoinformatics*, Wuhan, pp. 1–4, 2015.
16. N. Ojha and A. Vaish, "Spatio-temporal anomaly detection in crowd movement using SIFT", *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, pp. 646–654, 2018.
17. K. R. Kurte, S. S. Durbha, R. L. King, N. H. Younan and A. V. Potnis, "A spatio-temporal ontological model for flood disaster monitoring", *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX, 2017, pp. 5213–5216, 2017.
18. Peña-Araya, Vanessa, Mauricio Quezada, Barbara Poblete, and Denis Parra. "Gaining historical and international relations insights from social media: spatio-temporal real-world news analysis using Twitter", *EPJ Data Science*, Vol. 6, No. 1, 2017.
19. M. F. Bin Tarek, M. Asaduzzaman and M. Patwary, "Spatio- Temporal Analysis of Large Air Pollution Data", *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, Dhaka, Bangladesh, pp. 221–224, 2018.
20. R. Yadav and S. Kumari Sheoran, "Modified ARIMA Model for Improving Certainty in Spatio-Temporal Crime Event Prediction", *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*, Jaipur, India, pp. 1–4, 2018.
21. M. Beuchert, S. H. Jensen, O. A. Sheikh-Omar, M. B. Svendsen and B. Yang, "aSTEP: Aau's Spatio-TEmporal Data Analytics Platform", *2018 19th IEEE International Conference on Mobile Data Management (MDM)*, Aalborg, pp. 278–279, 2018.
22. Hao Fan, Chuanfeng Zhao and Yikun Yang, "A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018", *Atmospheric Environment*, Volume 220, 2020
23. G. Venkata Rao, K. Venkata Reddy, Raghavan Srinivasan, Venkataramana Sridhar, N.V. Umamahesh, et al., "Spatio-temporal analysis of rainfall extremes in the flood-prone Nagavali and Vamsadhara Basins in eastern India", *Weather and Climate Extremes*, Volume 29, 2020.
24. Gwitira, I., Mukonoweshuro, M., Mapako, G. et al. Spatial and spatio-temporal analysis of malaria cases in Zimbabwe. *Infectious Diseases of Poverty* 9, 146 (2020).
25. Alvaro Briz-Redon and Angel Serrano-Aroca, "A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain", *Science of The Total Environment*, Vol. 728, 2020.

Figures

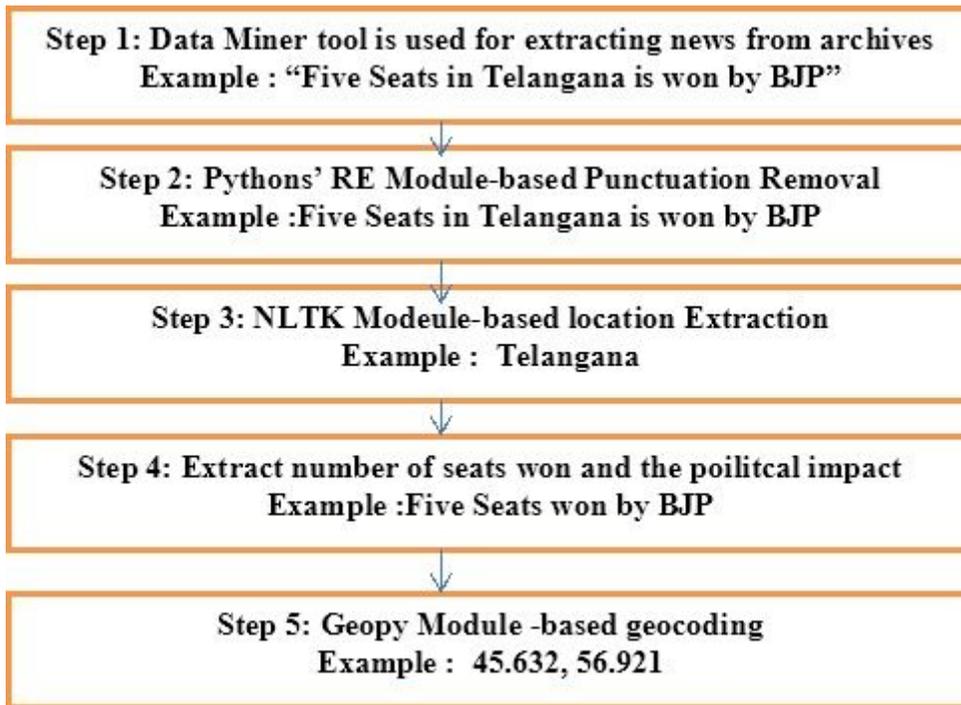


Figure 1

Steps involved in the proposed PPED-STAScheme

Vote share and seats of the National Democratic Alliance, Lok Sabha, 2019 elections

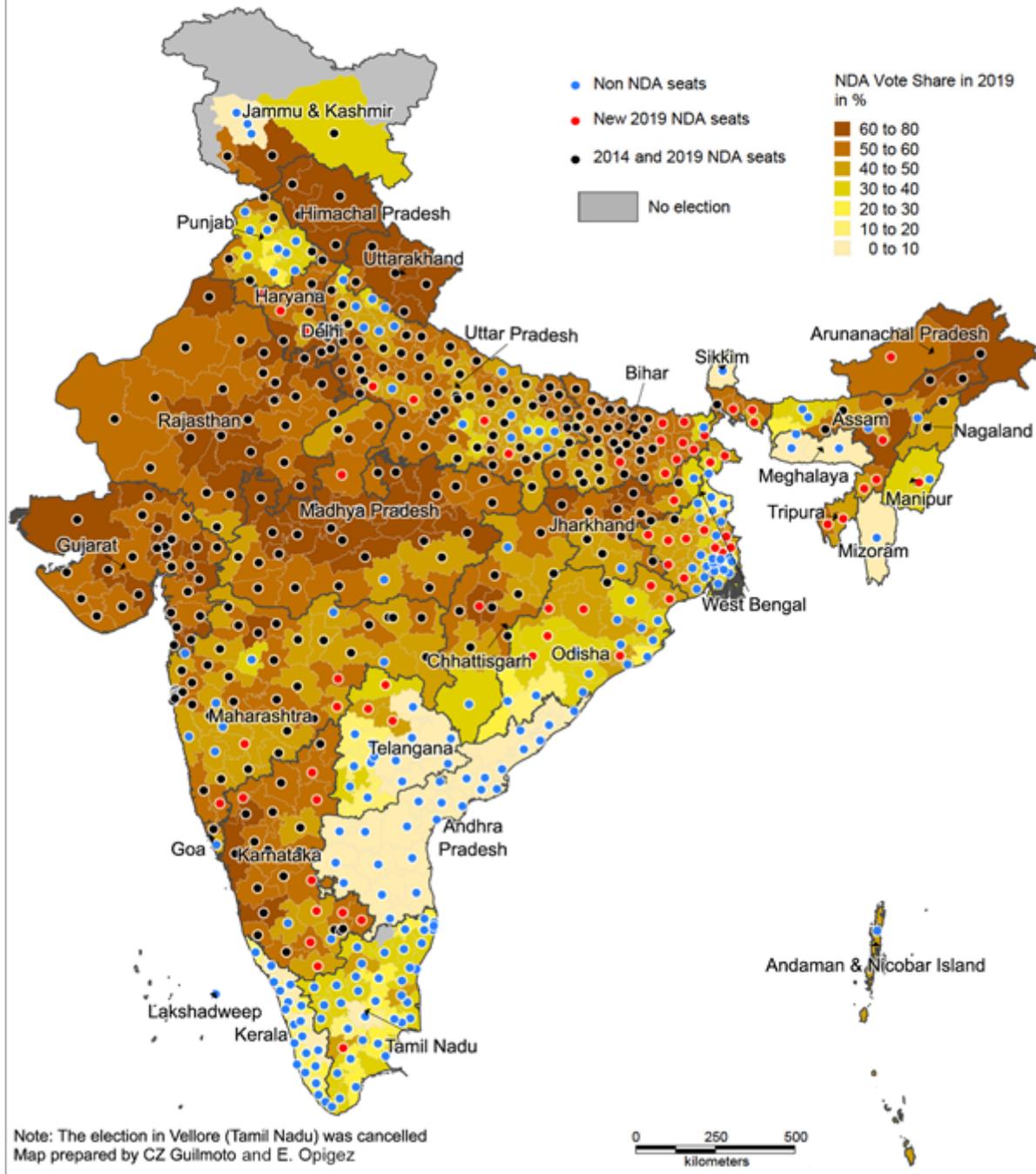


Figure 2

Vote Share and seats of National Democratic Alliance in 2019 elections.

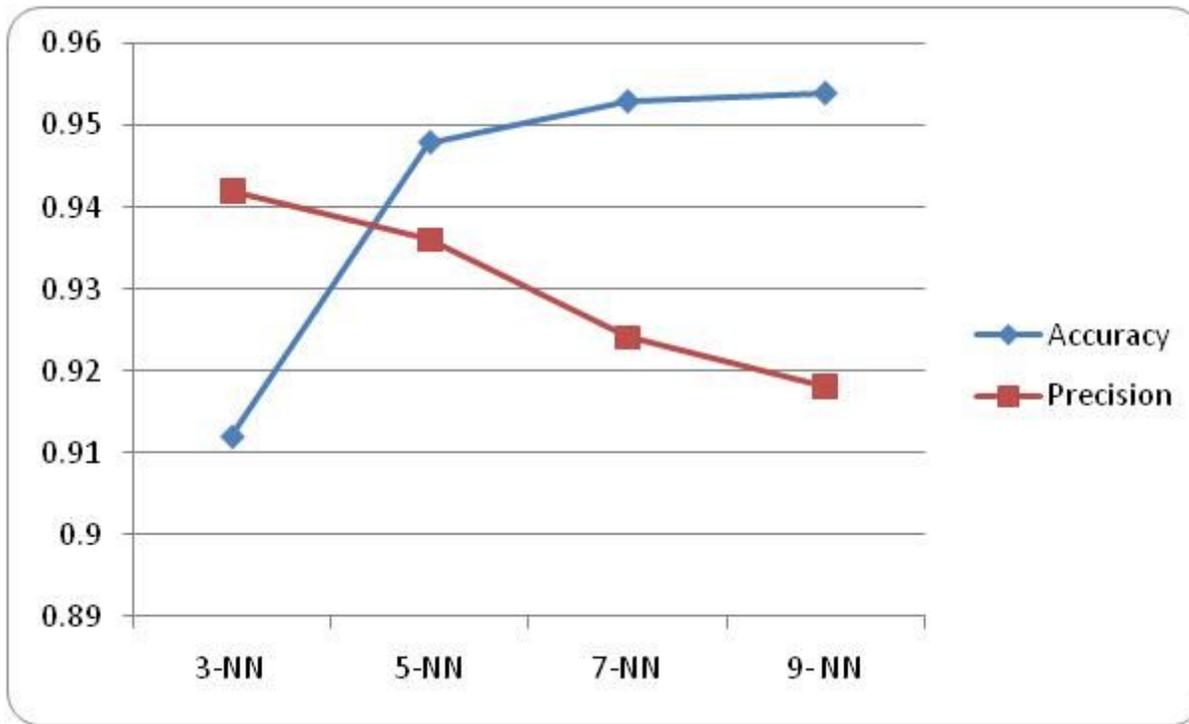


Figure 3

The proposed PPED-STA Accuracy and Precision based on k-NN attained

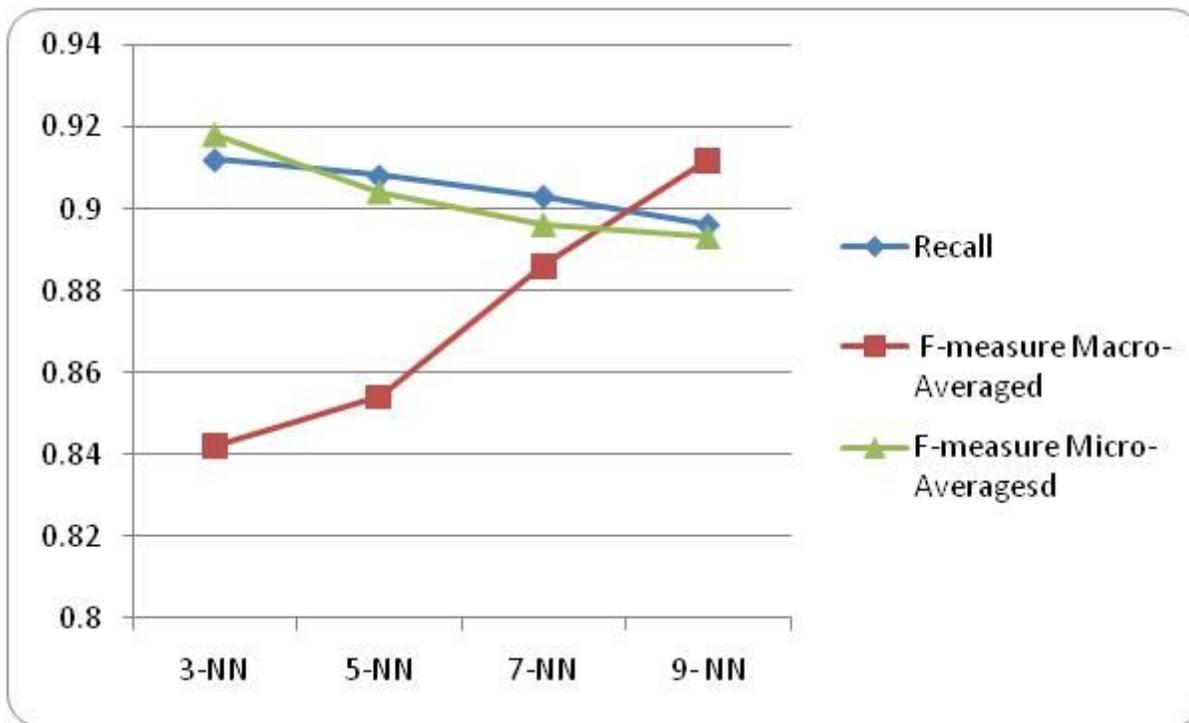


Figure 4

Proposed PPED-STA: Accuracy and Precision based on k-NN

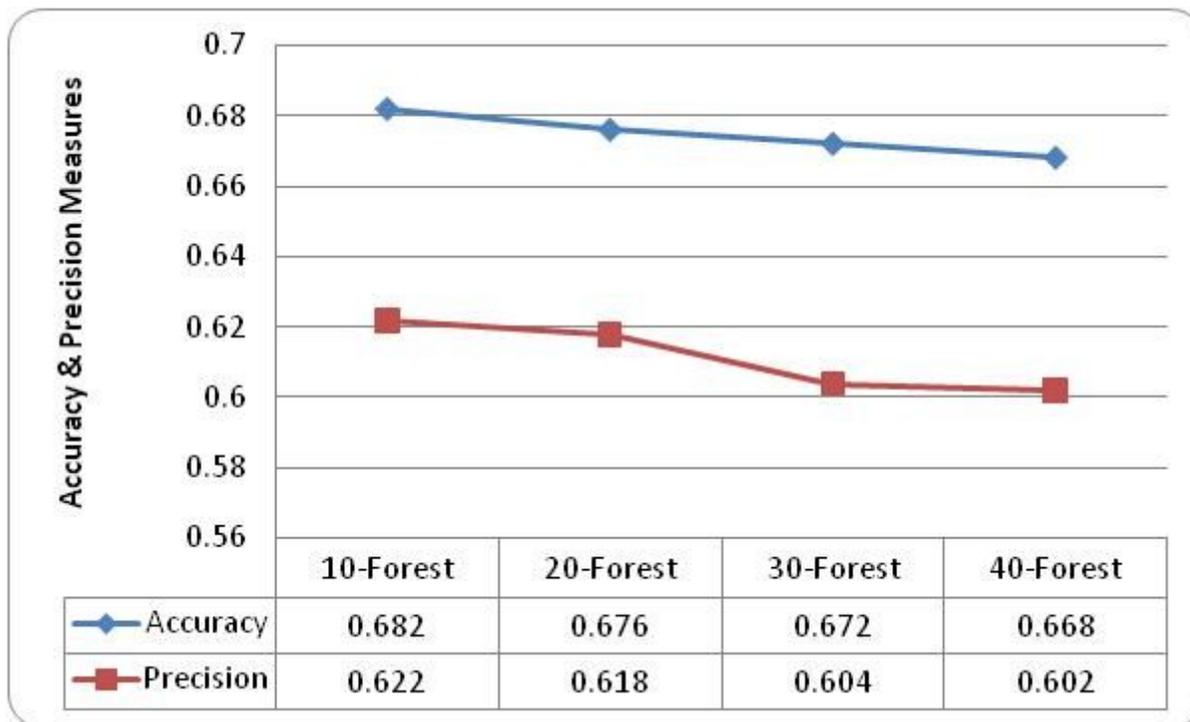


Figure 5

Proposed PPED-STA: Accuracy and Precision based on Random Forest

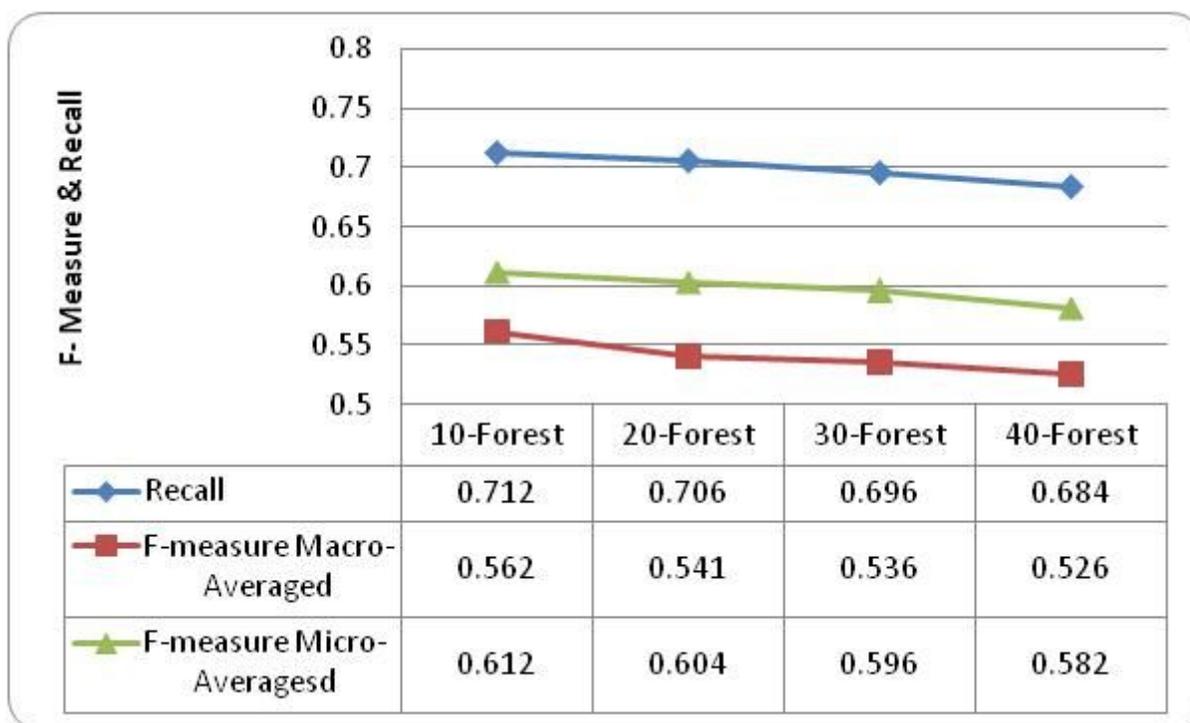


Figure 6

Proposed PPED-STA: Accuracy and Precision based on Random Forest

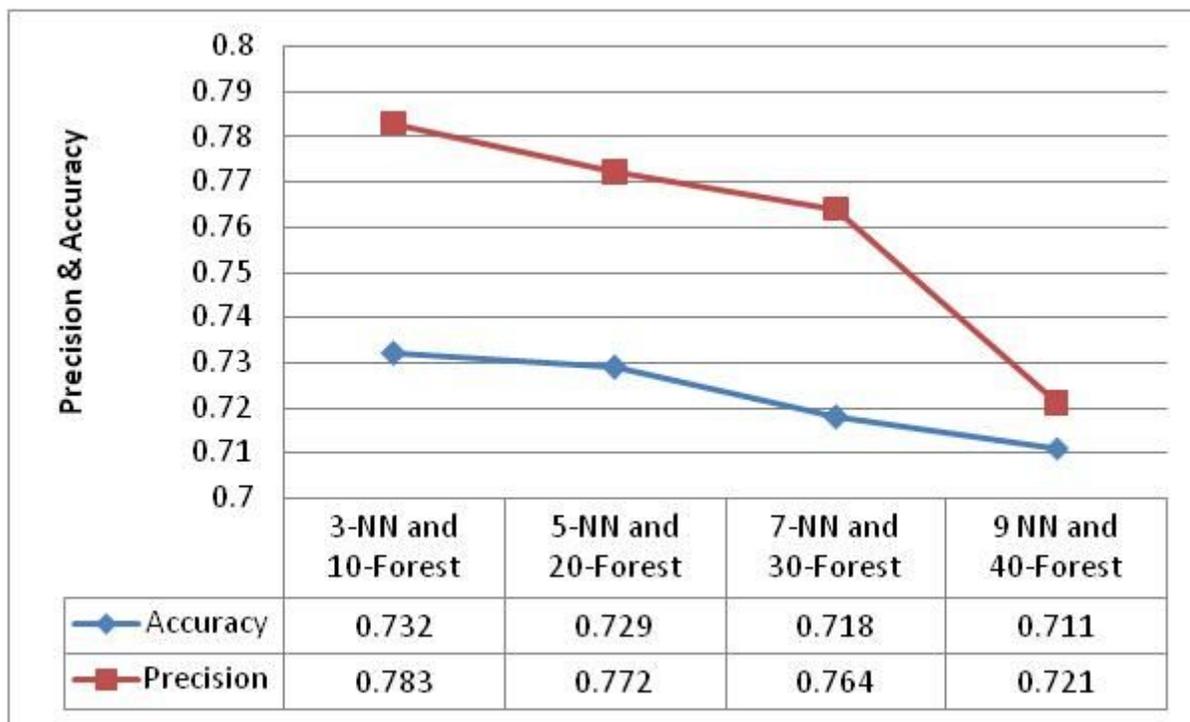


Figure 7

Proposed PPED-STA: Accuracy and Precision based on Random Forest

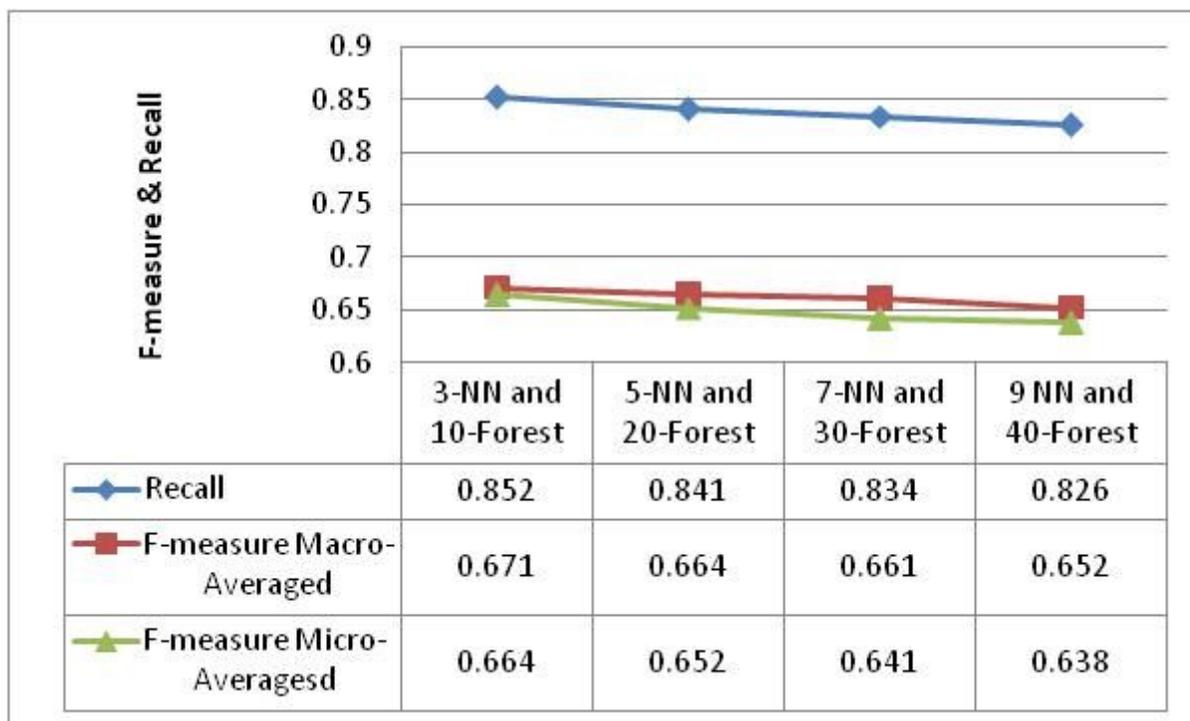


Figure 8

Proposed PPED-STA: Accuracy and Precision based on kNN and k-Random Forest