

# Enhancement of Classification using FPF-ANN for Big data Analysis in Distributed Environment

K Murali Gopal, Pragnyaban Mishra, R. P. Singh



**Abstract:** The development of massive amount of information from any source of group at any time, wherever and from any device which is termed as Big Data. The age group of big data becomes a dangerous challenge to grip, take out and access these data is short length of time. The detection of everyday itemsets is an significant issue of data mining which helps in engendering the qualitative information for the business insight and helps for the verdict makers. For the extracting the necessary itemsets from the big data a variety of big data logical techniques has been evolved such as relationship rule mining, genetic algorithm, mechanism learning, FP-growth algorithm etc. In this paper we suggest FP-ANN algorithm to promote the FP enlargement calculation with neural networks to maintain the feed forward approach. The recommend algorithm uses the Twitter social dataset for the collection of frequent itemsets and the proportional analysis of this approach is done using the different performance measuring parameters such as Precision, Recall, F-measure, Time complexity, Computation cost and time. The simulation of proposed work is done using the JDK, JavaBeans, and Wamp server software. The experimental results of projected algorithm gives better results in deference of time difficulty, computation cost and time also. It also gives enhanced results for the Precision, recall and F-measure.

**Keywords :** Big Data Analytic, Genetic Algorithm, FP-Growth, Association Rule, Neural Network, Precision, Recall, F-measure.

## I. INTRODUCTION

Big Data is a narrative term assigned to the datasets which emerge large in size; we cannot supervise them with the established data mining techniques and software tools available. “Big Data “appears as a tangible large size dataset which hides any information in its huge volume, which cannot be explored without using new algorithms or data mining techniques.[1] We have all heard of the 3Vs of big data which are Volume, Variety and Velocity, yet other Vs that IT, business and data scientists need to be worried with, most notably big data Veracity.

- **Data Volume:** Data volume gauges the measure of information open to an association, which doesn't unavoidably have to possess every last bit of it as stretched out as it can induction it.

As information volume increment, the estimation of disparate information records will reduce in relation to age, type, extravagance, and amount among different components.

- **Data Variety:** Data choice is a decide of the thriving of the information portrayal – content, pictures video, sound, and so forth. From an explanatory perspective, it is possibly the greatest check to productively utilizing extraordinary volumes of information. Jumbled information positions, neutral information structures, and clashing information semantics speaks to significant difficulties that can prompt logical spread.

- **Data Velocity:** It quantifies the speed of information creation, gushing, and accumulation. The web-based business has expediently enlarged the pace and lavishness of information utilized for exceptional business exchanges (for instance, site clicks). Information speed the board is considerably more than a data transmission issue; it is likewise an ingest issue.

- **Data Veracity:** Data veracity alludes to the inclinations, clamor and deviation in information. Is the information that is being put away, and mined huge to the difficulty being broke down. Veracity in information psychotherapy is the chief test when looks at to things like volume and velocity.[1]

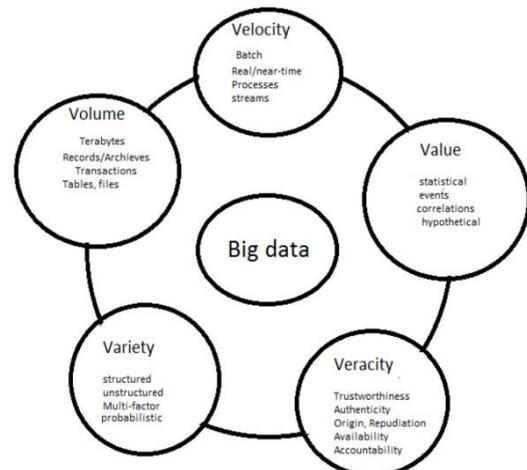


Fig 1.1: Five V's of big data [2]

### Big Data Steps for Processing

1. **Data Acquisition:** The first step in Big data is obtaining the data itself. With the mounting normal the pace of data age section is developing exponentially. With the presentation of exquisite resource which is utilized with a broad determination of sensors perpetually produce data. The Large Haudron Collider in Switzerland produces petabytes of data.

Revised Manuscript Received on June 30, 2020.

\* Correspondence Author

**K Murali Gopal\***, Computer Science and Engineering, Satya Sai University of Technical and Medical Science (SSSUTMS), Sehore, Madhya Pradesh, India. E-mail: kmgopal@giet.edu

**Dr. Pragnyaban Mishra**, Computer Science and engineering, KL University, Vijayawada, India E-mail: pragnyaban@gmail.com

**Dr. R. P. Singh**, Computer Science and Engineering, Satya Sai University of Technical and Medical Science (SSSUTMS), Sehore, Madhya Pradesh, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Enhancement of Classification using FPF-ANN for Big data Analysis in Distributed Environment

A large portion of this data isn't helpful and can be pointless, however because of its unstructured structure; specifically evacuation the data presents a test.

This data turns out to be progressively compelling in nature when it's joined with different inestimable information and superimpose. Because of the interconnectedness of gadgets over the World Wide Web, data is increasingly being gathered and put away in the cloud. [8]

2. Data Extraction: The entirety of the data created and gained isn't of utilize. It contains a lot of unnecessary or immaterial data. For instance, a basic CCTV camera, consistently surveys sensor to assemble data of the client's exercises. Nonetheless, when the client is in a state of fixed status, the data produced by the development sensor is neglected and of no utilization. The difficulties open in data withdrawal are twofold: above all else, because of landscape of data created, choosing which data to remain and which to discard constantly more relies upon the condition where the data was initially produced. For example, recording of a surveillance camera with similar edges might be unnecessary anyway it is basic not to dispose of practically identical data for a situation where it is being produced by a pulse sensor. Furthermore, an absence of an all inclusive stage presents it's have set of difficulties. Because of open assortment of data that exists, carrying them underneath a regular stage to standardize data taking out is a significant test.

3. Data Collation: Data from a solitary source regularly isn't sufficient for examination or forecast. More than one Data sources are frequently common to give a better picture than break down. For instance a state of being screen application much of the time gathers Data from the pulse sensor, pedometer, and so forth to condense the wellbeing data of the client. In like manner, atmosphere forecast programming takes in Data from various sources which make known the day by day soddenness, temperature, precipitation, and so on. In the plan of Big Data combination of Data to shape a superior picture is regularly viewed as an exceptionally huge piece of handling.

4. Data Structuring: When all the data is accumulated, it is noteworthy to close by and store data for extra use in a prearranged group. The organizing is basic so inquiries can be made on the data. Data organizing utilizes techniques for sorting out the data in a specific construction. Diverse new stages, for example, NoSQL, can inquiry even on unstructured data and are by and large always utilized for Big Data Analysis. A most significant issue with large data is given that constant outcome and in this manner organizing of accumulated data should be done at an expedient pace.

5. Data Visualization: When the data is organized, inquiries are finished on the data and the data is realistic in a visual arrangement. Data Analysis includes focusing on regions of interest and given that outcomes dependent on the data that has been organized. For instance, data containing standard temperatures are uncovered along the edge of water usage rates to assess a relative in the middle of them. This examination and game plan of data makes it arranged for utilization for clients. Crude data can't be utilized to expand bits of knowledge or for making a decision about examples, thus —humanizingl the data turns into even more significant.

6. Data Interpretation: The possible advance in Big Data agreement incorporates explanations and increasing costly

data from the data that is handled. The data picked up can be of two sorts:

i) Retrospective Analysis remembers for advanced bits of knowledge with respect to occasions and moves that have recently made spot. For the event, data about the TV viewership for a show is not at all like territories can assist us with umpiring the notoriety of the show in those regions.

ii) Prospective Analysis incorporates making a decision about models and insightful patterns for desires from data that is recently been created. Climate figure utilizing enormous data investigation is a case of anticipated examination. Issues collecting from such understandings are reasonable for fraudulent and equivocal patterns being anticipated. This is prevalently perilous because of developing a dependence on data for key choices. For instance, if an exacting manifestation is plotted close by the likelihood of being determined to have a careful illness, it may direct to purposeful publicity about the sign being caused because of the fastidious infection itself. Bits of knowledge picked up from data understanding are then extremely basic and the most significant explanation behind preparing enormous data also. All sections must be discouraged. All passages must be solid, for example, both left-legitimized and right-advocated. [8] We systematize relax of the paper in such a way: Section II presents the former work done by the diverse researchers for the study of the big data. Section III momentarily describes the assorted techniques of BDA. Section IV obtainable the anticipated work and Section V temporarily discusses concerning the untried system and their analysis. Last but not smallest amount presents the generally conclusion and future work of the proposed work.

## II. LITERATURE SURVEY

*Sanjay Rathee and ArtiKashyap (2018)* recommended an Adaptive-Miner: a resourceful dispersed association rule mining algorithm on Spark. The extraction of costly data from general datasets is a champion encircled by the most principal investigation issues. Affiliation rule mining is one of the amazingly utilized techniques for this guideline. Discovering the plausible relationships between things in enormous agreement based datasets (finding incessant itemsets) is the most basic piece of the affiliation rule mining task. For these sorts of equal/appropriated applications, MapReduce is extraordinary compared to other flaws in open-minded structures. Hadoop is one of the most acknowledged open-source programming systems with MapReduce based push toward for scattered extra room and giving out of incredible datasets utilizing detached groups worked from administration equipment. Versatile Miner is a powerful affiliation to decide mining calculation that adjusts its methodology dependent on the regular history of the dataset. Subsequently, it is disparate and improved than cutting edge static affiliation rule mining calculations. We disposition top to bottom investigations to expand understanding into the effectiveness, productivity, and versatility of the Adaptive-Miner calculation on Spark.[3] *Rochd and Hafidi (2018)* determine the trouble of behavior dataset in every iteration,

we nearby an algorithm called Hybrid Frequent Itemset Mining on Hadoop ( HFIMH ) which uses the perpendicular layout of dataset to resolve the difficulty of handling the dataset in every iteration. Vertical dataset conveys information to determine carry of each itemsets, and the thought of set crossroads is utilized to compute it. We weigh against the carrying out of HFIMH with one more Hadoop based completion of Apriori algorithm for dissimilar datasets. Experimental results reveal that our move toward is better.[4] *Dongmei Ai et al.,(2018)* Designed and algorithm Association rule mining algorithm on high-dimensional datasets. In this paper they expressed that the study of bioinformatics has been quickening at an expedient pace, presenting increasingly facial appearance and dealing with unrivaled volumes. however, these quick changes have, at the equivalent time, presented difficulties to information mining applications, in fastidious ingenious affiliation rule mining. Numerous information digging algorithms for high-dimensional datasets have been put advance, however the total quantities of these algorithms with episodic highlights and application situations experience issues settling on appropriate decisions. Therefore, we present a general study of different association rule mining algorithms suitable to high-dimensional datasets.[5]

*Francisco Padillo et al. (2019)*, built up that the proposition were keen to run on Big Data (document evaluates to 200 GBytes). The examination of uncommon quality measurements uncovered that no factual uniqueness can be set up for these two methodologies. At long last, three not at all like measurements (accelerate, scale-up, and size-up) have likewise been broke down to exhibit that the proposition scale in reality well on Big Data. Ends: The untried examination has uncovered that sequential calculations can't be utilized on huge amounts of information and approaches, for example, CBA-Spark, CBA-Flink, CPAR-Spark, or CPAR-Flink are necessary. CBA has end up being useful when the significant objective is to achieve exceptionally interpretable outcomes. Be that as it may, when the runtime must be limited CPAR ought to be utilized. No arithmetical dissimilarity could be built up between the two recommendations in specifications of the predominance of the outcomes separated from for the interpretability of the last classifiers, CBA being measurably better than CPAR.[6]

*Prajapati et al. (2017)*, anticipated strategy at first concentrates repetitive itemsets for each zone utilizing realistic scattered basic example mining calculations. The paper additionally matches the time the great association of Mapreduce based ordinary example mining calculation with Count Distribution Algorithm (CDA) and Fast Distributed Mining (FDM) calculations. The affiliation produced from visit itemsets is too incredible that it gets multifaceted to research it. In this way, Mapreduce based trustworthy and conflicting principle location (MR-CIRD) calculation is wanted to identify the reliable and inconsistent standards from large information and present valuable and noteworthy comprehension to the region specialists. These pruned intriguing principles additionally give helpful information for upgraded showcasing methodology also. The extricated consistent and conflicting guidelines are assessed and analyzed dependent on different intriguing quality activities open commonly with untried outcomes that lead to the finishing up ends. This is an open permission article under the CC BY-NC-ND permit.[7]

### III. BDA TECHNIQUES

For the examination of Big data different techniques has been urbanized such as association rule mining, inherited algorithm, appliance learning etc. In this segment, we are amplification some of them in brief.[9]

#### A. Association rule learning

Association rule (AR) learning is a strategy for finding fascinating association between variable with regards to extraordinary databases. It was essential utilized by most significant grocery store chains to discover intriguing relations between items, utilizing information from market retail location (POS) frameworks.

Association rule learning is animal used to help:

- Position items in better quickness to each extra so as to enlarge deals
- Extract data regarding the matter of guests to sites from web server logs
- Analyze hereditary information to uncover new connections
- Monitor framework logs to recognize interlopers and angry movement
- Recognize if individuals who purchase endeavor and margarine are increasingly likely to purchase diapers

#### B. Classification tree analysis

Statistical classification is a strategy of distinguishing classes that new observation has a place with. It requires a preparation set of suitably recognized perceptions recorded information at the end of the day. The factual arrangement is being utilized to:

- Automatically dole out archives to classes
- Classify life forms into groupings
- Develop profiles of understudies who take online courses

#### C. Genetic algorithms

Genetic Algorithms Are Stimulated By The Way Evolution Works – That Is, During Mechanisms Such As Birthright, Alteration And Ordinary Selection. These Mechanisms Are Used To "Develop" Useful Solutions To Problems That Require Optimization. Hereditary Algorithms Are Being Used To:

- Schedule specialists for medical clinic calamity rooms
- Return course of action of the most ideal materials and designing practices essential to create eco-friendly vehicles
- Generate "misleadingly inventive" placated, for example, plays on words and jokes

#### D. Machine Learning

Machine Learning incorporates programming that can concentrate from information. It gives computer the fitness to concentrate without being plainly customized, and is listening cautiously on making expectations support on realized property gained from sets of "training information." AI is being utilized to help:

- Discriminate among spam and non-spam email messages

# Enhancement of Classification using FPF-ANN for Big data Analysis in Distributed Environment

- Find out client inclinations and figure suggestions dependent on this data
- Determine the best fulfilled for connecting with forthcoming clients
- Determine the possibility of winning a case, and setting legitimate charging rates

## E. Regression Analysis

At a fundamental level, relapse examination includes controlling some self-ruling variable (for example ambient melodies) to perceive how it control a dependent lopsided (for example time spent available). It depicts how the estimation of a poor variable changes when the autonomous alterable is differed. It works best with unending quantitative information like weight, speed or age.

Deterioration analysis is being utilized to decide how:

- Levels of client satisfaction influence client dependability
- The number of supports calls customary might be inclined by the withstand figure given the first day
- Neighborhoods and size influence the list cost of houses
- To discover the adoration for your life by means of internet dating destinations.

## F. Sentiment Analysis

Sentiment analysis assists specialists with finishing up the slants of speakers or scholars with the concession to a subject. Sentiment investigation is being utilized to help:

- Get better help at an inn network by investigating organization remarks
- Customize motivating forces and administrations to address what clients are truly looking for trouble
- Conclude what purchasers really think dependent on sentiments from web-based social networking

## G. Social Network Analysis

Social network analysis is a system that was first utilized in the broadcast communications industry, and afterward quickly received by sociologists to contemplate relational connections. It is currently being utilitarian to break down the connections between individuals in numerous fields and attractive exercises. Hubs compare to people inside a system, while attaches compare to the connections between the people.

Social network analysis is being utilized to:

- See how masses from bizarre populaces structure ties with outcasts
- Find the criticalness or control of a specific character inside a gathering
- Find the most minimal measure of direct binds necessary to associate two people
- Understand the social structure of a buyer base

## IV. PROPOSED METHODOLOGY

### ASSOCIATION MINING PROPOSED WORK: EXTENDING FP APPROACH FOR ASSOCIATION MINING OVER SOCIAL DATA.

This section presents the Feed Forward Artificial Neural Network (FP-ANN) which calculates the hash sticker and

discovers significance between inputs. The planned FP-ANN upgrades FP development reckoning with neural networks to maintain the feed frontward come near. Key aspects of the projected FP-ANN are given under.

- Use of ANN Feed Forward Algorithm for the hash label and discovers the significance between the words input.
- In this network, the information moves in only one track, onward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.
- Commonly speaking, if one is prearranged a graph instead of a feed forward network, it can forever be grouped into layers such that each layer depends merely on layers to its left.
- This algorithm creates a confusion tag with high rate of accuracy and in adding up preferable outcome carrying out over previously utilizes commonplace FP-development algorithm for the hash label.
- An info dataset from the agreed micro blogging stage is taken and processed by diverse sub process library.
- FP-ANN, feed onward layer is based on ANN and it is utilized for calculations to linger productive while sentence the information and receiving hash label age greater than the extensive dataset. Figure 2 below gives an design how the proposed FP-ANN works.

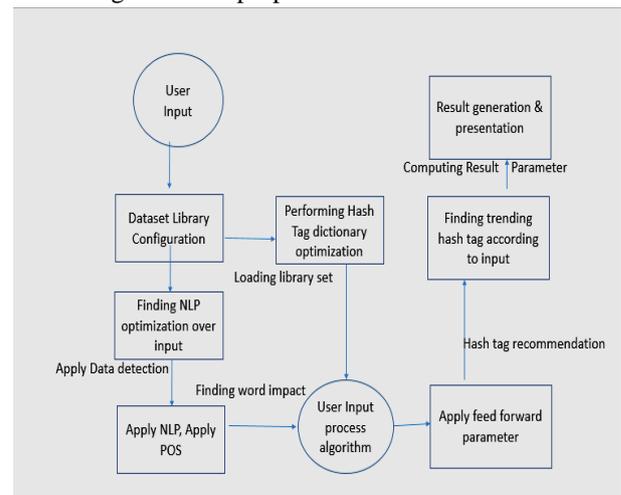


Fig. 4.1: Data Flow diagram of proposed work.

### Working steps of Proposed Algorithm:

- Step 1: In the original step the input will be taken from the user as per the require. Listing and loading of all the parameter, section for the replication purpose and relationship of the complete mandatory scenario framework.
- Step 2: In this step the arrangement of the documentation containing datasets will performed also hash tag dictionary optimization will be completed in this step. Listing and loading of all the parameter, constituent for the simulation purpose and configuration of the entire compulsory scenario framework.
- Step 3: This step will effort on finding the NLP optimization over input.
- Step 4: In this step the natural language processing will be applied over the given set of inputs.



Top HashTag		Bottom HashTag	
HashTag	Place	HashTag	Place
MyBrand	Mumbai, India	SLVIND https://t.co/ZASOhrH4	null
JUSTIN	New Delhi	WWC17	Dubai, United Arab Emirates
NDTVNewsBeeps	India	BREAKING	Indore
BREAKING	Indore	NDTVNewsBeeps	India
MakeInIndia	Bhopal	JUSTIN	New Delhi

## Top and Bottom Association Hash tag generation

So, we have to write code for twitter thoughts and then link our source code with Apache to get pleasure from the functionality of the twitter API.

- The presented algorithm take improvement over preceding traditional techniques but still more refinements are obligatory as per today's typical. Thus, an improved security, hashing mechanism can make it more consistent and executable to embark upon with present security and cloud scenario in the world.
- FP enlargement algorithm makes a repetitive computation and accurateness over the recurring value so that it is quite a flourishing algorithm in generating better outputs.

**Table 1:** Computation time comparison

Number of tweets (1K=1000)	Current State of art (ms)	FP-ANN(ms)
1K	14.36	12.2
2K	17.98	16.0
3K	23.12	20.9

## OVERHEAD COMPARISON

This experimentation is performed with the intention of calculating above your head in the proposed FP FF-ANN and in progress state of the art as transparency play detrimental effect in performance.

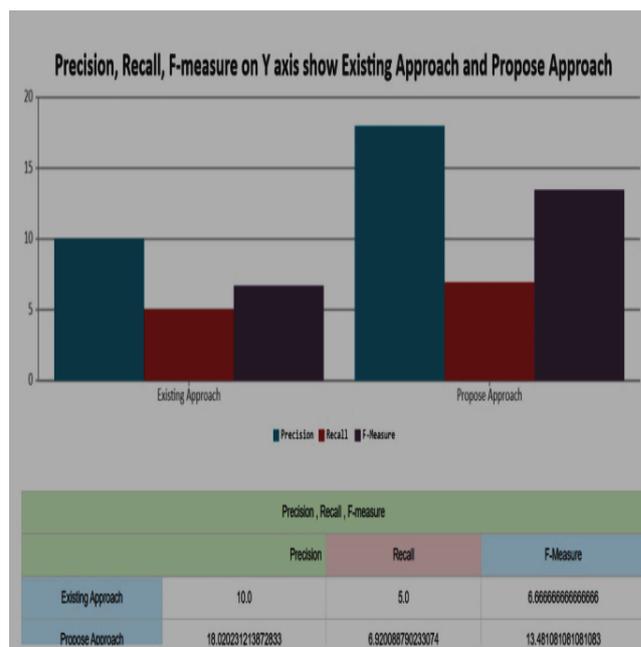
**Overhead comparison**

Number of tweets (1000=1K)	Current State of art	FPFF-ANN
1K	5454	4340
2K	7845	4908
3K	8081	7668

## B. RESULT ANALYSIS

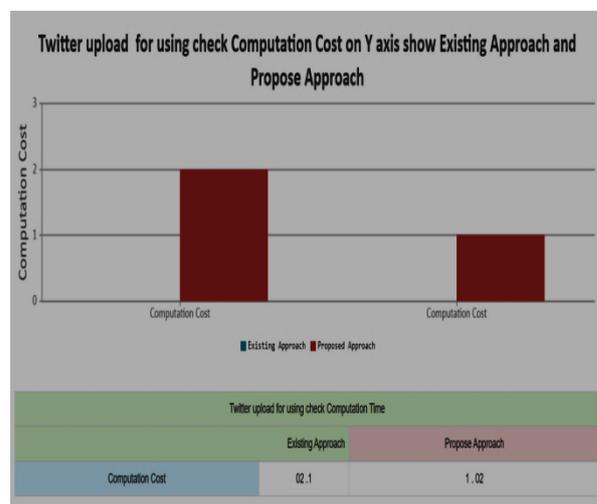
The analysis of proposed methodology is done using the Twitter dataset which is taken from the cancer forum site and this dataset enclose the subsequent field such as Twitter Stream, Hash Tag, Hash Count, Treand, NLP&POS, NLP&POSOLD FP-Growth, Search. We use Precision, Recall, F-measure, Computation Time, Computation Cost and Time Complication Parameter for the proportional analysis between the proposed methodology and existing methodology. The graph 5.1 shows the analysis of proposed methodology and exiting methodology using the exactness, call to mind and F-measure parameters and it is set up that the our proposed methodology gives enhanced results than the

existing come close to such as precision is 8% more, bring to mind about 2% more and F-measure is about 7% more.



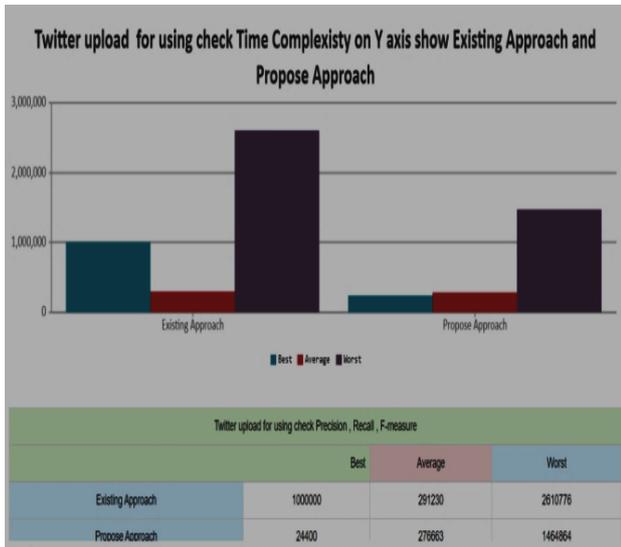
**Graph 5.1** Analysis of Twitter Dataset Using Precision, Recall and F-measure between Proposed and Existing Methodology

The graph 5.2 shows the analysis of proposed methodology and exiting methodology using the Computation cost parameters and it is originate that the our proposed methodology gives enhanced results than the existing come within reach of which is about 1 % less computational cost than the existing approach.



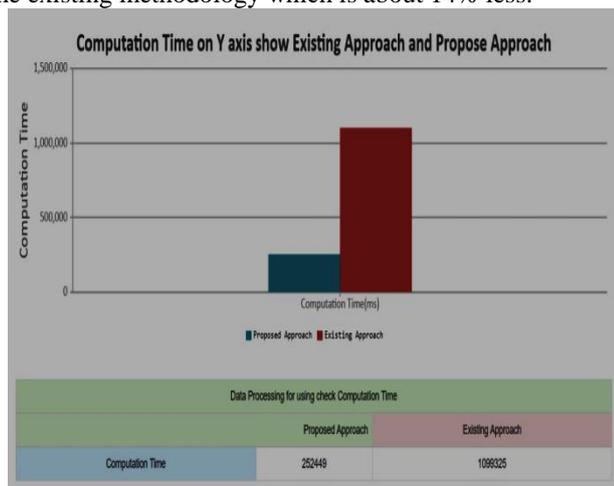
**Graph 5.2:** Analysis of Twitter Dataset Using Computation cost parameter between Proposed and Existing Methodology

The graph 5.3 shows the analysis of proposed methodology and exiting methodology using the Time convolution parameters and it is designed in all the best, most awful and standard cases. In all the cases of time complication it is found that our proposed methodology takes very less time in computation or in effecting of the algorithm than the existing methodology.



**Graph 5.3: Analysis of Twitter Dataset Using Time Complexity parameter between Proposed and Existing Methodology**

The graph 5.4 shows the investigation of proposed methodology and exiting methodology using the Computation time parameters which is a time compulsory to execute computation of the resources and it is found that our proposed methodology takes very less computation time than the existing methodology which is about 14% less.



**Graph 5.4: Analysis of Twitter Dataset Using Computation Time parameter between Proposed and Existing Methodology**

**VI. CONCLUSION**

All over the gigantic amount of data is together from the diverse source of purpose crossways a wide selection of area which is probable to double each couple of years. These data is positive only when we get indispensable information after analysis. For big data analysis it is requisite the expansion of techniques and this can be probable using a commanding computer realize these technique leads to mechanical system and for high presentation huge scale data dispensation is not trouble-free task to convert the data into knowledge by comprising exploiting parallelism of current computer architecture for data mining.

There are different data mining techniques has been residential for the analysis of big data such as support vector machine (SVM), neural network, fuzzy logics, irregular sets etc. which is also successful for the analysis . Big data are reduced to consist only those attribute of data which is imperative for the purpose area. So, for this diminution model has been implemented. In this work, we use an collection approach for the analysis of big data in disseminated environment is Feed Forward Artificial Neural Network (FPFF-ANN) which upgrades the FP-growth with neural network. This proposed work is protected and took less computation time and along these lines computational rate over the prepared available dataset. ANN in the FPFF-ANN enables it to execute quick calculations. It shows its helpfulness and calculates parallel measure. These calculations likewise check for apposite duplication, usage of more secure and trustworthy parameters. The analysis of projected work is done using Social media dataset (Twitter) and proportional analysis is execute with unusual measuring parameters such as Precision, Recall, F-measure, Computation cost, Time complication and Computation time with existing approach. The replication of propose work is done using poles apart software like Java development kit (JDK), NetBeans and WAMP server for the database. This work gives the superior results for the exactitude, recall and f-measure than existing approach is about 8%, 2%, and 7% more. Correspondingly the analysis of proposed work is done using computation cost parameter and it is originate that our work gives 1% more enhanced result than the existing approach. Later it is also compared with existing approach by time complication and computation time in which our proposed work reduces time complication and computation time regarding 14% than existing approach. On the whole it is analyze that our work is healthier in all reverence than the exiting approach which improves the effectiveness of the system successfully. In prospect work, our proposed method will be functional on additional data mining techniques such as clustering and classification because the basic phase of classification clustering entails finding everyday pattern associated to fastidious classes or clusters. Nevertheless, for more secure workload, additional research is desirable to determine a way of loading the input data to the nodes. More than that, to authorize even well again effecting times, a hybrid implementation can be achieved among in-memory and Java implementations. Also the implementation of proposed work is performing in several other parameters of like accurateness and specificity to get more accurate results of it.

**ACKNOWLEDGEMENTS**

We would like to express thanks the not mentioned referees of the paper for their expensive recommendations and productive comments which helped us in cleansing the quality of the paper.

**CONFLICT OF INTEREST**

The authors speak out that there are no conflicts of interest concerning the publication of this paper.



## REFERENCES

1. K.Arun and Dr.L.Jabasheela, "Big Data: Review, Classification and Analysis Survey", International Journal of Innovative Research in Information Security (IJIRIS) ISSN: 2349-7017(O) Volume 1 Issue 3 (September 2014).
2. SHERIN A et al., "SURVEY ON BIG DATA MINING PLATFORMS, ALGORITHMS AND CHALLENGES", International Journal of Computer Science & Engineering Technology (IJCSET) Vol. 5 No. 09 Sep 2014 ISSN : 2229-3345.
3. Sanjay Rathee and ArtiKashyap, "Adaptive-Miner: an efficient distributed association rule mining algorithm on Spark", <https://doi.org/10.1186/s40537-018-0112-0J> Big Data (2018).
4. Yassir ROCHD and Imad HAFIDI "An Enhanced Apriori Algorithm Using Hybrid Data Layout Based on Hadoop for Big Data Processing" IJCSNS International Journal of Computer Science and Network Security, VOL.18 No.6, June 2018.
5. Dongmei Ai et al., "Association rule mining algorithms on high-dimensional datasets" Artificial Life and Robotics (2018) 23:420–427.
6. Francisco Padillo et al "Evaluating associative classification algorithms for Big Data."Big Data Analytics (2019) <https://doi.org/10.1186/s41044-018-0039-7>.
7. Dinesh J. Prajapati et al., "Interesting Association Rule Mining with Consistent and Inconsistent Rule Detection from Big Sales Data in Distributed Environment" Future Computing and Informatics Journal (2017) <http://www.journals.elsevier.com/future-computing-and-informatics-journal>.
8. Subaira. A. S, Gayathri. R, Sindhuja a. N, " Security Issues and Challenges in Big Data Analysis", International Journal of Advanced Research in Computer Science and Software Engineering 6(2), February - 2016, pp. 530-536.
9. Debbie Stephenson "https://www.firmex.com/resources/blog/7-big-data-techniques-that-create-business-value/".