IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue-16, May 2019 E-ISSN: 2347-2693

# Fake News Detection: A Survey

**Divya[1*], N. Mehala[2]**

[1,2] Department of Computer Science, Presidency University, Bangalore , India

*Corresponding Author: divya@presidencyuniversity.in*

DOI: https://doi.org/10.26438/ijcse/v7si16.8187 | Available online at: www.ijcseonline.org

***Abstract***— Social media plays a vital role in online news circulation due to its ease of access, low cost, swift diffusion of information. The news or information can be of any topic, propagated in multiple modalities. Because of the huge amount of information exchange through online or through different social media platforms such as Facebook, twitter, differentiating true news from the low quality news that leads to the problem of fake news, which may affect the individual or the growth of the society, has become a challenging task. This calls for identification and filtering of fake news. In this paper our focus is on exploring the existing methods or approaches on content (i.e. text) based fake news identification, existing standard data sets, evaluation metric(s), tools and future scope, which helps the researchers to turn up with different and efficient approaches to identify fake news.

***Keywords***—*Fake news, Social networks, Multimedia*

## I. INTRODUCTION

Social Media is one of the most popular way /medium to gather companions and communities to share their views, ideas, experience and thoughts etc. The portraying sensations of the present-day occasions restyling the world as we undoubtedly aware, social media acts as a mode for the discussions, corporate schemes, photo sharing junctures, societal betting, microblogs, social broadcasting, talk applications in formal societies. Thus, social media have impacted on human conduct a lot further than other media and acts as one of the major source for the exponential development of data on the web. Furthermore, social media characteristics like its ease, simple access, and quick dispersal of facts that lead the people to look out and access news from web based life. On the other hand, it empowers the extensive spread of "forged newscast", i.e., news with deliberately untrue facts. Also, it is becoming exceedingly hard to separate reality from the false [3]. So, this prompts the issue of fake news. The wide-ranging spread of false news has the prospective for strong destructive effects on populations and society [5]. Fake news discovery has pulled in a developing enthusiasm from the overall population and scientists as the flow of deception online increments, especially in news sources, for example, internet based life channels, news sites, and online papers and so forth. Fake news identification presents special attributes and difficulties that make existing data recognition calculations from conventional data ineffectual or not relevant. Further, few significant difficulties/issues are identified from the past research papers: First, fake news is purposely composed to delude users to accept false data, which marks it troublesome and consequential to recognize using only available information; along with, we have to incorporate users' information, for example, client social commitment via web-based networking mass media, to help make a conviction [2 ]. Second, incorporating users' information increases the difficulty due to the nature like huge, deficient, amorphous, and noisy [1]. Different types of false news is condensed into various classes: **1. *Optical based*** false news which utilizes photographic representation of data. This incorporates utilization of photograph shopped pictures, videos, and blend of both images and videos [4], **2. *Client centered*** false news uses the gathering of people grouped based on their intended interest to speak about particular topic in which they are interested as business, culture, and so forth[8], **3. *Post-based:*** Post-based false news focuses on data collected via web-based networking media stages such as posts. Post can be a Facebook status or updates alongside picture or video and subtitle, a peep in twitter, image [7], **4. *System based/Network-based*** news are situated towards specific individuals from a particular association that are related in one or the other way, this is like the companions are connected on Facebook page and gathering of commonly associated people on LinkedIn [14], **5. *Erudition/Learning based*** false news comprise logical or sensible clarification to an uncertain disputes, these kind of newscast sections are intended to blowout deceitful data, e.g. untrue artefact on the most proficient method to fix asthma, **6. *Style-based*** false news concentrates on the approach of presenting to its consumers, as these false news are composed by larger part of individuals who are not professionals – so the style of writing can be unique and different from the professional writers and **7. *Position founded*** or ***Posture-built*** sort in-lines

 

with previously stated style-based sort, it centers on how explanations are given in an article. Because in a true artifact user can consume the detailed information but fake article comprises the statements which leads to arguments by providing insufficient information [24]. Authentic news stories are written in an approach to provide adequate data roughly on the topic and it is on per users to take way the importance of the story. Position founded stories are composed to give almost no data about the topic and to put forth a great deal of expressions (counterfeit arguments)[8]. Along these, we have directed this review to additionally encourage examine on this issue.

Fake news data can be of different forms: **1**. ***Content***: Text/string which mostly centers on the content as a correspondence framework. It is substantially more than simply sentence and words, it has qualities like tone, language structure, and has pragmatics that permits talk examination***, 2. Multimedia***: It is a combination of various types of media which incorporates pictures, video, sound, and designs. This is extremely visual and grabs watchers' eye at first [6,8], **3**. ***Audio***: Audio is a piece of mixed media class, however it has independent intermediate to be a newscast source. This class incorporates digital recording, communication range, radio administration. This medium connects with the more prominent group of listeners to convey the news [8] and **4**. ***Hyperlinks or Embedded Content***: Hyperlinks empower authors to embed the related information in the different forms such as Facebook post, tweet, YouTube video, sound cloud clip, Ingram post, etc. and gains users trust based on the embedded information.

In this paper, we have focused on few existing methodology for false news identification and depicting instantly about the strategies, informational indexes, data sets, evaluation metrics and future direction based on only the content based data. Web based life has ended up being an amazing hotspot for false news scattering. A survey on existing false news recognition techniques can give an essential comprehension on the best in class counterfeit news identification methods, give an insight of expanding the field to other applications and to develop a robust system [15].

## II. APPROACHES

A significant number of existing false news identification strategies have been proposed in the literature. In this segment, we discuss few methodologies and their basic characterization.

*1. Semantic Features based Methods/ Linguistic based methodologies:*
These methods are tied in with utilizing/removing key phonetic highlights from false news, for example, N-Grams, Punctuation, Psycho-Etymological highlights, Readability and Syntax etc [8].

- N-Grams: In the fields of computational historical underpinnings and probability, a n-gram is an abutting gathering of n items from a given case of substance or talk. Phonemes, syllables, letters, words or base sets can be the things as indicated by the application. The n-grams normally are amassed from a substance or talk corpus. Right when the things are words, n-grams may in like manner be called shingles [8][2][1][10].
- Punctuations: Letters, numbers and unique characters are doled out a sound, while accentuation checks and spaces are a piece of the structure of the composed dialect and are handled not through vocalization but rather are utilized to pass on structure [24].
- Psycholinguistics or brain science of dialect is the investigation of the interrelation between etymological components and mental aspects.[1] It likewise thinks about mental and neurobiological factors that empower people to procure, use, appreciate and deliver dialect. The order is predominantly worried about the components in which dialects are handled and spoke to in the mind [34].

*2. Deception Modeling based Methods:*
The way toward grouping beguiling versus honest stories depends on hypothetical methodologies: Rhetorical Structure Theory (RST) and Vector Space Modeling (VSM)[13]. This procedure includes applying RST, which results in each dissected content that is changed over to a lot of expository relations associated in a progressive tree, VSM is then used to distinguish the consequences of logical structure associations [10].

*3. Grouping grounded Approaches:*
Grouping is a realized strategy to thoroughly analyze a lot of information, in[6], creators have utilized gCLUTO (Graphical CLUstering TOolkit) bunching bundle to support separate newscast information dependent on their similitude dependent on picked bunching calculation. This strategy includes running an extensive number of informational index and framing/arranging few bunches utilizing agglomerative grouping with the k-closest neighbor approach, grouping comparative news reports dependent on the standardized recurrence of associations.

*4. Prescient Demonstrating based Methods:*
In these methods, creators proposed a calculated relapse demonstrate dependent on preparing informational collection of 100 out of 132 news reports. As indicated by this methodology, positive coefficients increment the likelihood of truth and negative one increment the likelihood of duplicity. This technique gives 70% of precision on

 

preparing informational index and 56% of exactness on test information set[6, 7].

*5. Content Cues based Methods:*
This strategy use two unique examinations:

- Lexical and Semantic Levels of Analysis: Selection of vocabulary expect a fundamental employment in convincing perusers to place stock in the section. Computerized procedures can be used to isolate stylometric features of the substance (i.e., linguistic frame, word length and conceptual terms) that can be used to isolate among two article introductions. [8]
- Syntactic and Logical Planes of Study: Logical capacity of features summons reference to approaching portions in the talk [7]. This is finished by constructing orientation to imminent chunks in the news section. Features are composed to fill void considerations with utilizing resulting content. This investigation likewise covers estimating news locales which have more offer action contrasted with destinations that considerably creates more news content [8].

*6. Non-Text Cues based Methods:*
A news require not to be content compulsorily, it very well may be picture or video, so this technique investigations the news in two diverse ways for example Picture examination and client conduct investigation.

## III. DATA SETS

Much logical research depends on the social occasion and examination of estimation information. Logical informational indexes are, at any rate, intermediate results in numerous logical research ventures. For quite a while informational indexes weren't distributed and regardless of whether they were distributed it was for the most part as a (not re-usable) side-effect of the production. Be that as it may, an intriguing marvel may be seen here: informational collections (regularly in mix with models and parameters) are ending up increasingly imperative themselves and can at times be viewed as the essential scholarly yield of the research [31]. At the point when the distinctive creators proposed diverse methodologies for false news recognition, these proposed methodologies are estimated or tried with various classes of informational collections, as it is one of the vital factor to confirm the work. News can be gathered from various sources, for example, news organization, landing pages, web crawlers, web based life, publicly supporting, day by day schedule investigation and sites. These informational collections can be extensively characterized into Open informational collections which is accessible for the client at no expense and Commercial Data sets which is accessible to the client or customer at some expense. As indicated by this study the most regularly utilized informational indexes are BuzzFeedNews [1][8], Liar [1][8], BS locator[1], Credbank[1][9], Politifact[5][6], PHEME[8] and publicly supporting and so forth. These informational collections are again ordered into various classes in another way for example Web-based social networking based and API based[4][6][7]. Web based life based informational indexes are Facebook remarks, post and the tweets in twitter etc. API based informational indexes are information gotten by reality checkers, for example, BS Detector, Politifact and Routine every day news and so forth. Informational indexes are likewise arranged into various classes dependent on their sources, for example, News article informational collections, story based, Website, Journalism and Social media stages. In few existing researchers' work, informational indexes can be considered as manually marked and checked informational indexes.

## IV. EVALUATION METRICS

Measurements are parameters or proportions of quantitative appraisal utilized for estimation, assess, correlation, or to follow execution or generation. Measurements are quantitative estimates intended to help assess look into yields. Measurements can't give a basic response to complex inquiries. As a proportion of consideration, measurements can just inform you so much regarding the quality, achievement or effect of research and specialists. Numerous creators utilize distinctive measurements to assess the precision of their proposed methodologies. The most regularly utilized measurements are Precision and Recall, F1-Score measure, ROC, Macro accuracy, Macro Recall, Macro F Score measure, Train and split model and so forth. In example acknowledgment, data recovery and parallel arrangement, accuracy is the division of significant cases among the recovered cases, while review is the percentage of important occurrences that have been mended over the collective sum of pertinent occasions. Both exactness and review are hence founded on a comprehension and proportion of significance. It has been observed that different evaluation metrics can be proposed by different researchers with respect to the context of application to assess the execution of the proposed methodology (ies).

- Precision and recall are defined as [32]:

$$Precision = \frac{tp}{tp+fp}, Recall = tp/(tp + fn) \quad (1)$$

Where tp: True Positive, fp: False Positive and fn: False Negative

- F1 Score is the weighted normal of Precision and Recall. This score considers both false positives and false negatives. Instinctively it isn't as straightforward as precision, however F1 is normally more helpful than exactness, particularly when we have an uneven class dissemination. Exactness works best if false positives and false negatives have comparative expense [35].

 

The F1 score is the consonant typical of the exactness and survey, where a F1 score accomplishes its best a motivating force at 1 (flawless exactness and review) and most exceedingly terrible at 0 [33].

F-score (F1 score) is the symphonious mean of exactness and review:

$F1=(recall^{-1}+precision^{-1})/2$

$F1=2\times(precision\times recall)/(precision+recall)$ (2)

- A beneficiary working trademark twist, i.e., ROC twist/curve is used to measure the performance of approaches, it is made by plotting the authentic positive rate (TPR) against the false positive rate (FPR) at various edge settings. The authentic positive rate is generally called affectability, audit or probability of revelation [4] in machine learning. The false-positive rate is generally gotten the drop out or probability of false alarm [4] and can be resolved as (1 − distinction).It can be thought as Power as a part of the Type I Error of the verdict rule (when the execution is resolved from just a case of the people, it might be thought of as estimators of these sums).

$$TPR = \frac{TP}{P} = \frac{TP}{TP+FN} = 1 - FNR$$

$$FPR = \frac{FP}{N} = \frac{FP}{FP+TN} = 1 - TNR$$

$$TNR = \frac{TN}{N} = \frac{TN}{TN+FP} = 1 - FPR$$

(3)

Where, TPR-authentic positive rate, FPR-false positive rate and TNR-genuine negative rate,

As we are centered on substance/text based discovery strategies, the study of few existing papers with respect to the context on Fake news recognition has been outlined in the Table 1 with various methodologies connected, informational collections utilized and with degree for the future work.

## V. CONCLUSION AND FUTURE SCOPE

Due to the increase in the popularity of the social media, the amount of information shared among multiple social media users is growing exponentially which in turn leads to the spreading of fake news. Fake news has solid negative effects on each customers and progressively broad society which made the researchers to discover approaches to recognize false data in this blasting, counterfeit news-plagued space. In this paper, we have investigated the issues in handling the false news by surveying the existing research works in terms of approaches, tools, data sets, evaluation metrics and future scope to empower the exploration network and specialists to take up the challenge of handling and identification of fake news effectively and efficiently with the advent increase in data size.

## APPENDIX

| Paper | Author(s) | Approach(s) | Data Set | Outcome | Evaluation metric | Future scope |
|---|---|---|---|---|---|---|
| Fake News Detection on Social Media: A Data Mining Perspective [1] | Kai Shuy, Amy Slivaz, Suhang Wangy, Jiliang Tang and Huan Liuy | Linguistic based, User based and Post based- Model construction such as news content model and social content model | 1.BuzzFeedNews[15]: This dataset contains an entire test of news distributed in Facebook from 9 news Offices over seven days near the 2016 U.S.<br>2. LIAR [ 17]: This dataset is collected from fact verifying API Politifact<br>3. BS Detector [12 ]: This dataset is gathered from a program expansion called BS locator created for checking news veracity.<br>4. CREDBANK[19]: it is a substantial scale publicly supported dataset of around 60 million tweets | Could recognize the false news dependent on portrayal and feature extraction | In this paper the author has used a metric to compare true with a negative cases, where true is, when a news is found to be really fake or really genuine as per [23 ]. | Approach can be stretched out to Data arranged, Model situated, application situated false news discovery rather than highlight arranged. |
| Detecting Fake | Monther | Bayes Net, Logistics | Manually Prepared | Appropriate just | Accuracy and | This approach can |

 

| | | | | | | |
|---|---|---|---|---|---|---|
| News in Social Media Networks [2] | Aldwairi, Ali Alwahedi | and Naïve Bayes classifier using WEKA | Dataset has been used. | for false misleading content discovery | Recall, F-Measure and ROC. | be reached out for different wellsprings of false news proliferation. |
| Fake news detection in social media [3] | Kelly Stahl | Linguistic Cue and Network Analysis approaches in which Naïve Bayes classifier, Support Vector Machines, and semantic analysis algorithms are used | - | Ready to distinguish printed or word based false news | - | This approach an be reached out for different types of false news rather literary frame alone |
| Online Fake News Detection Algorithm [4] | Sakeena M. Sirajudeen, Nur Fatihah A. Azmi, Adamu I. Abubakar | Designing of conceptual framework, Formulation of Algorithms for Detecting the source of fake news by confining IP address, and finding the state of the fake news and filtering and Development of a proof | Routine every day news[29] | Presents calculations and a evidence for an idea for recognizing and Sifting of operational false news. | - | Calculations can be developed to maintain a strategic distance from the proliferation of false news and sifted in like manner |
| Fake News Detection with Deep Diffusive Network Model [5] | Jiawei Zhang, Limeng Cui, Yanjie Fu, Fisher B. Gouza1 | Fake Detector, Deep Walk, Line, Propagation, Rnn and SVM | The dataset used in this work is combination of both data from an API Politifact which are tweets from twitter, and besides the fact check articles made concerning these announcements in the PolitiFact site[33] | Proposed show can recognize the false news articles, makers and Subjects in the system. | In the assessment, we will cast the believability derivation issue into a parallel class order and a Multi-class order issue individually, execution can be assessed by measurements, similar to Exactness, Macro Precision, Macro Recall and Macro F1 exclusively. | Spam Detection Research and Applications, Deep Learning Research and Applications |
| 3HAN: A Deep Neural Network for Fake News Detection [6] | Sneha Singhania, Nigel Fernandez, and Shrisha Rao | 3HAN-graded attention network-Stanford CoreNLP, Encoders are used to process sentences of bodies and tokenized sentences and features into words | PolitiFact [14] a regarded actuality checking site discharged a rundown of locales physically researched what's more, marked. Utilized those locales from this rundown named counterfeit | displayed 3HAN which makes news vector, a successful portrayal of an article for discovery as false news | used a train, endorsement and achieved test split of 20% \| 10% \| 70% for neural prototypes likewise, a train and test split of 30% \| 70% for word check grounded models | web application subject to 3HAN which gives acknowledgment of fake news as an organization and learns in a veritable time online path from new physically truth checked articles |
| "The Pope Has a New Baby!" Fake News Detection Using Deep Learning [7] | Samir Bajaj | Logistic Regression, Two-layer Feed forward Neural Network, Recurrent Neural Network, Gated Recurrent Units and Convolutional Neural Network with Max Pooling | Data drawn from two interesting sources, both without trying to hide zone. Data collected here is from Kaggale data base –which contains around 13000 articles on various subjects | The model could separate whether the news is false or genuine with content component | Precision model | For false news discovery, included highlights later on can be the wellspring of the news, including any related URLs, the point, distributing medium (blog, print, web-based social networking), |

 

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | nation or geographic locale of root, production year, just as phonetic highlights |
| Media-Rich Fake News Detection: A Survey [8] | Shivam B. Parikh and Pradeep K. Atrey | Linguistic Features based Methods, Deception Modeling based Methods, Clustering based Methods, Predictive Modeling based Methods, Content Cues based Methods and Non-Text Cues based Methods | Buzz Feed News is an accumulation of title and connections to a genuine description or a post that is thought about false news[26], LIAR is a seat checking system made accessible by University of California, Santa Barbara scientists12], PHEME This informational index incorporates talk tweets, gathered and explained inside the news coverage use instance of the task[21] and CREDBANK The main informational collection has contained online life information and enables clients to perform investigation on Twitter information[15] | portrayal of news story in the current diaspora joined with the differential substance types of news story and its effect on perusers. Accordingly, counterfeit news recognition approaches that are vigorously in light of content based investigation | - | Multi-modal Data-set, Multi-modal Verification Method, Source Verification |
| FakeNewsTracker: A Tool for Fake News Collection, Detection, and Visualization [9] | Kai Shu, Deepak Mahudeswaran, and Huan Liu | Counterfeit News Assembly, News Discovery and Fake News Picturing Social Article Fusion (SAF) model- Deep learning-based solution and Linguistic features | Gathered the information from the false checking sites in gushing way and look in Tweets for any social commitment identified with the news pieces. | A framework FakeNews Tracker, which gives general answers for information accumulation, intelligent representation, and expository displaying towards false news recognition. | Precision, Recall and F measure | Approach can be reached out for different highlights accessible in the dataset like top picks, retweets and informal community and learn highlights for the false news identification. structure could be reached out to distinguish counterfeit news continuously as it is executed in gushing way |
| Automatic Detection of Fake News [10] | Veronica Perez-Rosas, Bennett Kleinberg, Alexandra Lefevre Rada Mihalcea | Linguistic Feature Extraction, Linguistic Inquiry and Word Count software (LIWC) | One dataset is gathered by means of Publicly supporting covering six news areas (e.g., business, instruction). The second dataset is acquired Specifically from the web and covers big name news. | directed a set of learning analyses to assemble precise false news indicators, and accomplished correctness's of up to 76% | Precision, Recall and F score measure | Future work can incorporate meta highlights (e.g., number of connections to and from an article, remarks on the article), highlights from various modalities (e.g., the visual cosmetics of a site utilizing PC vision approaches) |

## REFERENCES


[1] Kai Shuy, Amy Slivaz, Suhang Wangy, Jiliang Tang, and Huan Liu " *Fake News Detection on Social Media:A Data Mining Perspective"*.

[2] Monther Aldwairi, Ali Alwahedi *"Detecting Fake News in Social Media Networks",**2018.***

[3] Kelly Stahl *" Fake news detection in social media"* **15 May 2018**.

[4] Sakeena m. Sirajudeen, nur fatihah a. Azmi, adamu i. Abubakar, *"online fake news detection algorithm"* **2005**.

[5] Jiawei Zhang1, Limeng Cui2, Yanjie Fu3, Fisher B. Gouza *"Fake News Detection with Deep Diffusive Network Model"*.

[6] Sneha Singhania(B), Nigel Fernandez, and Shrisha Rao *"3HAN: A Deep Neural Network for Fake News Detection"* **2017**.

[7] Samir Bajaj *"The Pope Has a New Baby!"* Fake News Detection Using Deep Learning, **2017**.


 

[8] Shivam B. Parikh and Pradeep K. Atrey *"Media-Rich Fake News Detection: A Survey"* **2017**.
[9] Kai Shu, Deepak Mahudeswaran, and Huan Liu *" FakeNewsTracker: A Tool for Fake News Collection, Detection, and Visualization"*.
[10] Ver´onica P´erez-Rosas1, Bennett Kleinberg2, Alexandra Lefevre1 Rada Mihalcea *"Automatic Detection of Fake News".*
[11] Greatmoon hoax. https://en.wikipedia.org/wiki/Great_Moon_Hoax. [Online; accessed **25-September-2017**].
[12] S. Adali, T. Liu, and M. Magdon-Ismail. Optimal link bombs are uncoordinated. In AIRWeb, **2005**.
[13] L. Akoglu, R. Chandy, and C. Faloutsos. Opinion fraud detection in online reviews by network effects. In ICWSM, **2013**.
[14] H. Allcott and M. Gentzkow. Social media and fake news in the 2016 election. Journal of Economic Perspectives, **2017**.
[15] E. Arisoy, T. Sainath, B. Kingsbury, and B. Ramabhadran. Deep neural network language models. In WLM, **2012**.
[16] K. Bharat and M. Henzinger. Improved algorithms for topic distillation in a hyperlinked environment. In SIGIR, **1998**.
[17] C. Castillo, D. Donato, A. Gionis, V. Murdock, and F. Silvestri. Know your neighbors: web spam detection using the web topology. In SIGIR, **2007**.
[18] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, **2001.**
[19] K. Chellapilla and D. Chickering. Improving cloaking detection using search query popularity and monetizability. In AIRWeb, **2006**.
[20] E. Convey. Porn sneaks way back on web. The Boston Herald, **1996**.
[21] H. Allcott, and M. Gentzkow. *"Social media and fake news in the 2016 election"*. Journal of Economic Perspectives, **Vol. 31, No. 2, 2017, pp. 211–236.**
[22] V.S. Foster, Vincent S. *"The Great Moon Hoax."* In Modern Mysteries of the Moon, Springer International Publishing, **2016, pp. 11-44**.
[23] v. pickard, *"media failures in the age of trump."* the political economy of communication **vol. 4, no.2, 2017, pp. 118- 122**.g. lee.
[24] Bernard e m jones *"Exploring the role of punctuation in parsing natural text"* .
[25] http://www.journalism.org/2016/05/26/news-use-acrosssocial-media-platforms-2016/
[26] http://www.bbc.com/news/uk-36528256
[27] https://en.wikipedia.org/wiki/Pizzagate conspiracy theory
[28] https://www.buzzfeed.com/craigsilverman/viralfake-election-news-outperformed-real-news-onfacebook?utm term=.nrg0WA1VP0#.gjJyKapW5y
[29] http://time.com/4783932/inside-russia-social-media-waramerica/
[30] https://www.nytimes.com/2016/11/28/opinion/fakenews-and-the-internet-shell-game.html? r=0
[31] http://lit.eecs.umich.edu/downloads.html.
[32] Ronald Dekker *"The importance of having data-sets"* **2006**.
[33] https://en.wikipedia.org/wiki/Precision_and_recall.
[34] https://en.wikipedia.org/wiki/Psycholinguistics.
[35] https://stackoverflow.com/questions/45963174/what-is-f1-score-and-what-its-value-indicates.

## Authors Profile

N. Mehala is an Associate Professor in the Department of Computer Science at Presidency University, Bengaluru, India. Before that, she was working as the faculty in the department of Computer Science and Information Systems (CSIS) at Birla Institute of Technology and Science Pilani (BITS-Pilani) Pilani campus, from 2006 to 2015. She has totally 17 years of teaching experience. Her research and teaching interests include web mining and search engine algorithms. Presently she is guiding three Ph.D. scholars. She has supervised more than 20 M.Tech Thesis and Dissertations. She was 1. Nucleus Member of WILP in BITS-Pilani. 2. Member of organizing and technical committee in "International Workshop on Data Analytics and Applications" (IWDAA 2013) organized by departments of Computer Science & Information Systems of Pilani & Goa campuses of BITS-Pilani from 26th Feb. to 1st Mar. 2013 at the K K Birla Goa Campus, Goa, India. 3. Member of organizing committee in TFIR-BITS Workshop on "Introduction to Graph and Geometric Algorithms, BITS-Pilani, Jan 22-24, 2009.

Ms. Divya has 3.6 years of Experience in teaching. She is Post Graduate from RVCE Bangalore M.Tech in Computer Science and Engineering. She has enrolled for Ph.D in the area of Data Mining at Presidency University. Her career objective is to grow personally along with the organization or institution that offers her a consistently positive atmosphere to learn new technologies and implement them for the betterment of the profession. Published papers in reputed conferences and journals. Teaching areas are Computer graphics, System software, Embedded systems, Software Testing and Programming languages. She has Secured Best Student Award, Has been placed First Grade in "The State Level Language Program" and Has Passed the State Level 'Personality Development – Jeevana Shikshana Contest' conducted by Vivekananda Yuva Vedhike.

 