International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
Article Received: *25 April 2026* ***Revised:*** *12 May 2026* ***Accepted:*** *30 June 2026*

An Analytical Study of Data Mining Techniques Applied to Social Media Data

Dr. K. Sathishkumar1,

Assistant Professor, Erode Arts and Science College, Erode, Tamilnadu. E-mail id :sathishmscgasc@gmail.com

Mrs. R. Prema2,

Research Scholar, Erode Arts and Science College, Erode, Tamilnadu. E-mail id-premanilani@gmail.com

ABSTRACT
Neurons in Artificial Neural Networks (ANNs) are highly interconnected components that collaborate to solve particular issues. ANNs have recently been used in voice and picture recognition. Character and pattern recognition using data modeling and statistical analysis to address forecasting and classification issues. The learning process of a neural network is the main topic of this study.

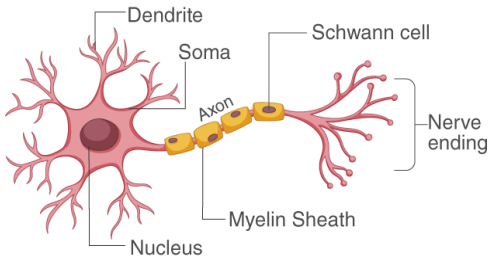
Keywords: supervised and unsupervised learning, ANNs, neurons, and pattern recognition.

I. INTRODUCTION

In biology, where neural networks are crucial to the human body, the idea of artificial neural networks (ANN) is essentially introduced. Neural networks are used to perform tasks in the human body. A neural network is just a network of millions upon millions of interconnected neurons. The human body is the best illustration of parallel processing as all of it is carried out by these interconnected neurons. [2]

II. HOW THE HUMAN BRAIN LEARNS

A typical neuron in the human brain uses a variety of tiny structures called dendrites to gather signals from neighboring neurons. Through a long, thin structure called an axon, which divides into thousands of branches, the neuron transmits electrical activity spikes. The activity from the axon is transformed into electrical effects that either inhibit or excite activity in the linked neurons at the end of each branch via a structure known as a synapse. A neuron emits a spike of electrical activity down its axon when it gets excitatory input that is large enough in comparison to its inhibitory input.



III. Artificial neural networks: what are they?

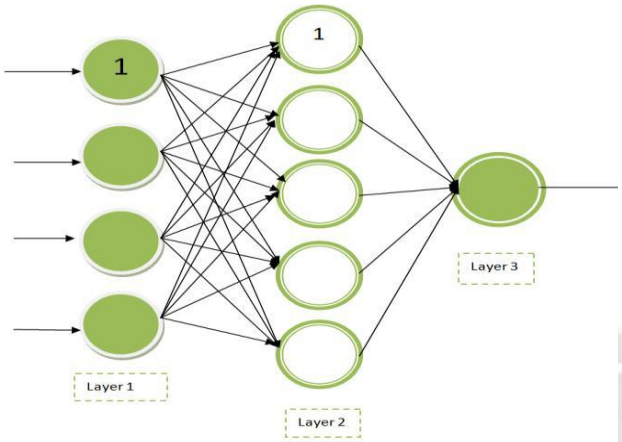
The way biological nervous systems, like the brain, process information serves as the model for an artificial neural network (ANN), a paradigm for information processing. The input layer, hidden layer, and output layer are the three layers that make up a neural network.

Information that can be stated numerically is received by the input nodes. The data is displayed as activation values, where a number is assigned to each node; the higher the number, the stronger the activation. After then, this data is distributed over the network. The activation value is sent from node to node based on the connection strengths (weights), inhibition or excitement, and transfer functions. After adding up the activation values it receives, each node adjusts the value according to its transfer function.

The activation travels via hidden layers in the network before arriving at the output nodes.

The input is then meaningfully reflected to the outside world by the output nodes.[4]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
Article Received: *25 April 2026* ***Revised:*** *12 May 2026* ***Accepted:*** *30 June 2026*

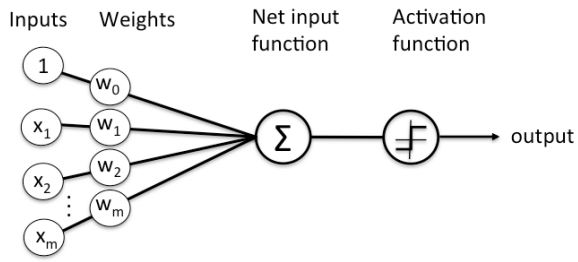


Why ANN?

The ANN has given the human brain several superior qualities that are absent from contemporary computers, such as: [2]

1)Capacity to Learn
2)Capacity for Generalization
3)Flexible Education
4)Tolerance for Faults
5)Organizing on Your Own

This flow diagram shows the active nodes that are utilized in the neural network's output and hidden layers. After multiplying each input by a weight, the total is calculated. This results in a single value that is run via an activation function.



IV. NEURAL NETWORK TYPES

Neural networks come in a variety of forms, but they are often divided into feed-forward and feed-back networks. With inputs, outputs, and hidden layers, a feed-forward network is a non-recurrent network where signals can only move in one direction.

A layer of processing components receives the input data and uses it to make calculations. The computation of each processing element is based on the weighted sum of its inputs. The subsequent layer is fed with the newly computed data. This procedure keeps going until all of the layers have been completed and the output has been determined. Sometimes a neuron's output in the output layer is measured using a threshold transfer function. Perceptrons, both linear and non-linear, and Radial Basis Function networks are examples of feed-forward networks. Data mining frequently makes use of feed-forward networks. [4]

Because a feed-back network features feed-back routes, signals can use loops to move in both directions. Neurons are permitted to form any kind of connection. This kind of network becomes a nonlinear dynamic system that undergoes constant change until it achieves a state of equilibrium because loops are present.

In associative memories and optimization problems, feed-back networks are frequently employed to find the optimal configuration of interconnected elements.[4]

The Process of Learning:

Neural network learning techniques can be divided into two main groups:

Supervised learning:

In this case, the target or desired pattern is linked to each input pattern used to train the network. When comparing the computed output of the network with the correct expected output to identify the inaccuracy, it is assumed that a teacher is present during the learning process. Performance can then be improved by using the error to modify network parameters.

[5]

Unsupervised learning:

The network is not shown the intended result while using this learning technique. The system learns on its own by identifying and adjusting to structural elements in the input patterns as it is as though there is no teacher to offer the required patterns. [5]

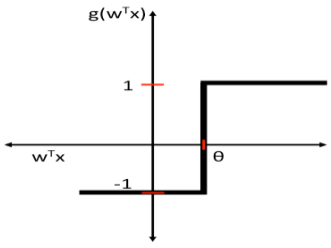
V. TRANSFER (ACTIVATION) FUNCTIONS

Both the weights and the input-output function (transfer function) that are provided for the units determine how an ANN (Artificial Neural Network) behaves. There are four common forms of transfer functions: Gaussian, sigmoid, piecewise linear, and unit step (threshold).

1) Unit step (threshold):

Depending on whether the total input exceeds or falls below a certain threshold value, the output is set at one of two levels.[4]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
Article Received: *25 April 2026* ***Revised:*** *12 May 2026* ***Accepted:*** *30 June 2026*

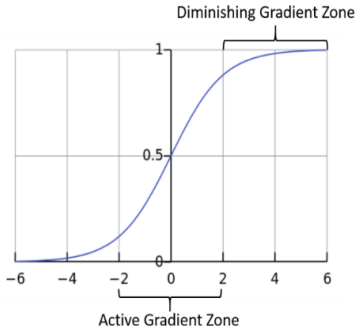


Unit step function.

2) Sigmoid

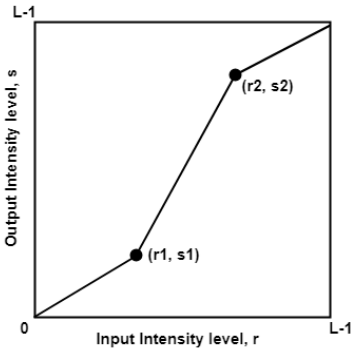
The logistic and tangential functions make up the sigmoid function. The logistic function's values fall between 0 and 1, while the tangential function's values fall between -1 and +1.[4]

$$A = \frac{1}{1+e^{-x}}$$



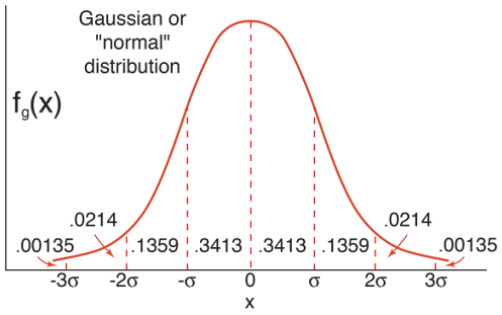
3) Linear Piecewise:

The output is proportionate to the overall output that has been weighted.[4].



4) Gaussian

Gaussian functions are continuous, bell-shaped curves. Depending on how near the net input is to a selected average value, the node output (high/low) is interpreted in terms of class membership (1/0). [5]



The following process can be used to teach a three-layer network to carry out a certain task: 1. We give the network training examples that include both the intended pattern of activities for the output units and a pattern of activities for the input units.

2. We assess the degree to which the network's actual output resembles the intended output. 3. To improve the network's approximation of the intended output, we alter the weight of each connection.

VI. NEURAL APPLICATIONS NETWORKS

Image (data) Compression: Data can be compressed and decompressed in real time using neural networks. Eight bits of data are reduced to three bits and then reversed back to eight

bits via the

Identifying Patterns:

A back-proliferation neural system that can distinguish between a real and a fake heart attack, a framework that can identify bombs in baggage at airport terminals by recognizing small differences and examples from inside specific sensor's yields, a system that can examine and further read the PAP smears, and many other examples of acknowledgment applications are in use. There are now a lot of robotized quality control applications that rely on pattern recognition.

Processing Languages:

Text-to-speech conversion, machine acoustic input, automatic language translation, secure voice-keyed locks, automatic transcription, voice-activated aids for the physically impaired and deaf, and natural language processing are some of these uses.

Character Identification:

Through the use of neural network-based products, hand-printed characters can be recognized by a scanner. For numerals, it is 97% accurate, and for alphabetic characters, it can reach 92%. Cursive characters can be recognized using the Quantum Neural Network software package **Processing Signals:**

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
Article Received: *25 April 2026* ***Revised:*** *12 May 2026* ***Accepted:*** *30 June 2026*

Electronic noise can also be filtered out by neural networks.

Financial: Banks, credit card firms, and lending organizations must make a lot of difficult decisions. Learning and statistical patterns are involved. These days, decision-making uses neural networks that have been trained using data from previous decisions.

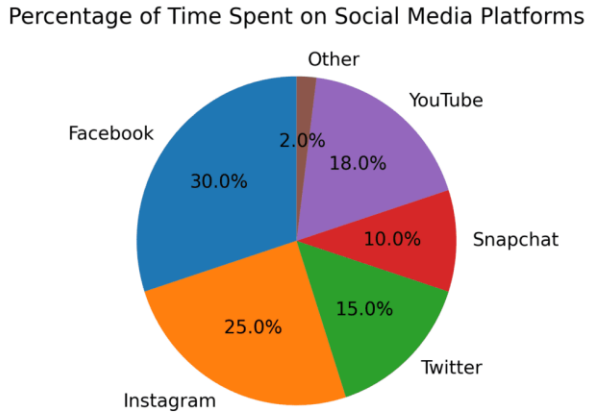
Social media is a remarkable means of societal communication. People can easily interact in the online world by sharing, uploading, commenting, blogging, liking, and tweeting information in many formats (text, audio, video, photographs, emails, multimedia, etc.) at any time and from any location. The analysis of social media data has become extremely important for analysts due to a number of factors. One multidisciplinary area of computer science is data mining. In data mining, we take information out of a massive data set and utilize the appropriate methods to identify patterns and knowledge for future use. Numerous allied domains, such as statistics, machine learning, pattern recognition, database systems, visualization, data warehouses, and information retrieval, depend heavily on data mining [1].

VII. THE FEATURES OF SOCIAL MEDIA

The term "social media" refers to a collection of web-based programs that enable the production and sharing of user-generated content and build upon the technological and ideological underpinnings of Web 2.0.[2] It offers people in the community a simple means of communication and a network on an unprecedented size and at speeds never found in traditional media. Wikis, social news, microblogging, media sharing, and other technologies have all evolved as a result of the widespread use of social media. The following table lists various social media platforms, their attributes, and the kinds of data those platforms use (see Table 1 below).

Benefit	Description
Customer Engagement	Real-time interaction fosters loyalty and understanding.
Marketing and Advertising	Targeted campaigns reach specific demographics effectively.
Lead Generation	Attracts potential customers and drives traffic to websites.
Brand Building	Establishes identity and expertise through consistent content.
Cost-Effectiveness	Affordable marketing compared to traditional methods.
Competitive Advantage	Enables small businesses to compete with larger firms.
Data Insights	Provides analytics for informed decision-making.
Crisis Management	Helps manage reputation and address issues swiftly.

Social Media Data: This data can be gathered with the help of traditional networks and social networks. Traditional networks include newspaper, News Channels, Movies, Radio, TV, FAX and television, etc. While social network sites are discussed in the above table 1. The popularity of social media continues to grow exponentially, in which user are added at each second; following pie chart(Figure 1) represents active users (in millions) on different social networking sites as of 2017. As we can see quantity of active users, thereby the data generated by the active users is also huge. Some of the statistics of data generation on these sites is shown in Table 2:



International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 2
Article Received: *25 April 2026* ***Revised:*** *12 May 2026* ***Accepted:*** *30 June 2026*

Comparison Table of Social Media Platform Users

This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

	Facebook	YouTube	LinkedIn	Twitter
User 1	20%	Your text here	Value	Value
User 2	Your text here	7%	20%	32%
User 3	5%	Value	Your text here	Value
User 4	Your text here	15%	Your text here	18%
User 5	Value	Your text here	25%	Your text here
User 6	10%	10%	10%	Value
	Conclusion	Conclusion	Conclusion	Conclusion

VIII. DIFFICULTIES WITH SOCIAL MEDIA DATA

Unfortunately, because social media data is collected from many social media platforms and in various formats, it differs from traditional statistics. In addition to its enormous size, it faces other unique difficulties, which are explained below:

Big Data: There is no doubt that social media data is enormous. A small piece of information about each person shows whenever we zoom in on them.

Acquiring Sufficient Samples: Social media sites' application programming interfaces (APIs) can be used to gather data. There is just so much data that can be collected each day. How can one rely on samples as well?

How can we make sure that the patterns we identify are real and useful for further research or business development?

Noise Removal Fallacy: Any irrelevant posts, tweets, messages, or other content are considered noise in social media data. However, eliminating the noise may render the data unusable.

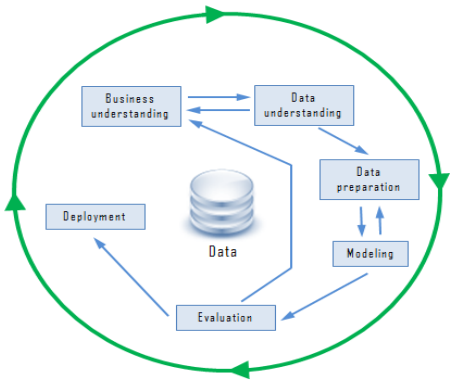
due to social media data's intrinsic connectivity. Compared to attribute-value data, it makes noise removal more difficult and necessitates a different method.

Evaluation Dilemma: A piece of the dataset is frequently used for testing in traditional data mining. However, typical test data might not be feasible with social media data. Analyzing social media data for patterns can provide an almost insurmountable problem because we cannot ensure the authenticity of the patterns without reliable evaluation.

Unstructured: Social media content created by users is frequently very unstructured. These days, a growing number of people use their phones to publish content, such as updating Facebook statuses, sending tweets on Twitter, and leaving comments on posts. This leads to the frequent occurrence of first short texts, second typos, and spacing errors [5]. **Incomplete:** The characteristics of users are predictable based on their personal information [6]. Social media companies frequently let their users utilize their profile settings to make their personal information—such as demographic profiles, status updates, friend lists, videos, images, and interactions on posts—invisible to others in order to allay such privacy concerns. **Missing Values:** The values that are absent in some cases. For instance, people might refrain from entering their age, location, or hobbies in their social media profiles. We can either (i) eliminate instances with missing values, (ii) guess missing values, or (iii) ignore missing values in order to fix this issue. examples that deviate significantly from other examples in the dataset are known as outliers. when calculating a person's likes on any given post. In that instance, a celebrity will be seen as an outlier since he can receive the most likes due to his popularity relative to

We must pre-process social media data because of its nature.

Data mining: Finding useful information in vast amounts of data is known as data mining. It is the analysis phase of the KDD, or knowledge discovery in databases, process.[7] It explains the standard procedure for deriving meaningful information from unprocessed data. Pre-processing, data mining, and post-processing are the three main components of the KDD process.



Pre-processing Social Data: Prior to using data mining techniques, we must pre-process the data after resolving the issues. The following are typical data pre-processing tasks: Aggregation is the process of combining several features into a single feature or altering a feature's scale. The process of converting quantitative data into qualitative data is known as discretization.

Feature Selection: Since most of the characteristics are useless due to their low processing power or lack of relevance, a selection is made that will improve the data mining technique's performance.

Feature Extraction: Unlike feature selection, feature extraction creates a new set of features that are better suited to the data mining activity [8]. A new set of features is extracted from the data after a transformation.

Data mining Methods: At this point, the data can be subjected to the proper methods. Supervised learning and unsupervised learning are two categories of techniques. A given data set is usually split into training and testing data sets with known class labels in supervised learning. [9] Using the training data, supervised algorithms create models that are then used to make predictions. It can be separated into two categories: regression and classification.

IX. CATEGORIES

This data is classified using a classification model into many classifications. It is referred to as classification when the class attribute is discrete. Decision trees, Naive Bayes classifiers, and K-nearest neighbors are popular classification techniques. **Decision tree:** It consists of nodes that form a directed tree, with the root node having no incoming edges. Review sets of positive and negative are utilized as leaf nodes in the approach. The most important review served as the tree's base.

Large data sets are employed with the Naive Bayes classifier. To apply conditional probabilities, Naïve Bayers counts the occurrences of values and combinations of values in historical data. It is thought to be one of the top three supervised learning strategies.[10] **K-nearest neighbor:** It does classification by using an instance's neighbors. Specifically, it performs classification using the k nearest examples, or neighbors. We must calculate an instance's distance from every other instance using a distance metric in order to identify its neighbors. Euclidean distance is frequently used.

Regression: Rather of being discrete, the item to be predicted in regression is continuous.

CONCLUSION

The intricate information concealed in social media data is the main topic of this paper. One can streamline his labor to produce the required data and reliable performance with the aid of the aforementioned strategies. For their future study in this field, this publication encourages and motivates scholars to review the numerous data mining methodologies.

REFERENCES

[1]. J. Han, M. Kamber, and J. Pei. Data Mining: Concepts and Techniques. Kaufmann, San Francisco, 2011. Morgan

[2]. A. M. Kaplan and M. Haenlein. Users of the world, unite! The challenges and opportunities of social media. Business Horizons 53(1):59{68, 2010

[3].https://www.statista.com/statistics/272014/glo bal-social-networks-ranked-by-number-of users/ visited on 28/2/2016

[4]. http://wersm.com/how-much-data-is generated-every-minute-on-social-media/ visited on 28/2/2016

[5]. D. Kim, D. Kim, E. Hwang, and S. Rho. Twitter trends: a spatio-temporal trend detection and recommendation Systems,2014. related scheme. keywords Multimedia

[6]. M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. PNAS, 2013.

[7]. Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases" (PDF). Retrieved 17 December 2008

[8]. Social Media Mining: An Introduction By Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, Pg No:142

[9]. Mining Social Media: A Brief Introduction: Pritam GundechaHuan Liu

[10]. Adedoyin-Olowe, M., Gaber, M. M., & Stahl, F. (2013). A survey of data mining techniques