Journal of Advances and Scholarly Researches in Allied Education
Vol. 20, Issue No. 4, October-2023, ISSN 2230-7540

A machine learning methods for forecast prediction on social media users

Renu Kumari[1]*, Vijay Kumar[2]

[1,2] College of Commerce Arts and Science, Patliputra University, Patna, 800020 Bihar, India

Abstract - The paper deals with "A machine learning methods for forecast prediction on social media users." The networking data is a very important aspect in the present and future networking data plays a very valuable role in decision making for social media user, organization, education sector, and service. With the online existence and social media users utilize various social media platforms like Instagram/Facebook or others app to express or comments their observations and opinions. The article network is a unweighted and directed network. Article-author network is an unweighted, undirected, and two-mode network. Author network is an weighted and undirected network. Machine learning works by progressively improving the performance of a given task. Previously, machine learning provides us with powerful web search techniques, speech recognition, home automation, automated surveillance system, self-driving cars, and much-improved perception about the genome.

Keywords - Machine Learning Approach, Vector Machine, and Social media User.

----------X----------

INTRODUCTION

Social network is an important tools for the accomplishment of the recently. Social network is broadly applied in a variety of applications and in many disciplines. It is being realized as the defacto tools for communication support. The details of modern social network analysis advances can be found in social media moods. However, collecting relational data from structured based documents, modelling networks, and obtaining actionable insights need expertizing and awareness in certain fields. ML of network analysis methods were developed to obtain knowledge from a variety of social networks. These methods were developed to examine connected/disconnected, weighted/unweighted, directed/undirected, bipartite, or labeled networks. The role of social media platforms are predicting Government Initiatives, Election results, product analysis, business analysis, movie popularity, sports outcomes and stock market analysis. This review paper proposed the opinions are expressed through different social media platforms can be used for retrieving or extracting the real time predictions on several trends. As per the sentiment identification outcome find the features in the form of positive-negative and Neutral. In this proposed research methodology, here collect users reviews on particular trends, then pre-processed it, creation of the features and selecting for data classification using different machine learning classifiers and predict the result. For better performance, used advanced preprocessing techniques will be applied to cleaning the data. For Sentiment Classification will be used machine learning algorithms or techniques as Support Vector machine, Maximum Entropy. Naive Bayes and Decision tree. As per existing techniques. It is very difficult to mine the correct predictions from social media. Therefore, the prediction model will be designed for doing the prediction using real time data from Instagram, Facebook and Twitter etc. An opinion from text or comment posted on social media platforms by various categories of users is one of the critical and time consuming tasks in the field of opining mining and analysis. The importance of this proposed intelligent system for social media is to automatically providing polarity from unstructured data for effective decision making. Machine learning is becoming an indispensable tool for automated knowledge discovery from historical data. Utilizing statistical techniques, machine learning allows computer to learn from data without being explicitly programmed.

Social networks: It is possible to identify significant actors, recognize communities, network propagation modelling, link/attribute prediction, entity detection, user behaviour analysis, community maintained resource support, study of geospatial interactions, analyse social distribution and filtering, designing of recommender systems, and several more. Therefore, it is essential to aggregate the techniques to construct a complete life cycle of such analysis. So we have carried out a study by analysing published research contributions related to betweenness centrality algorithm in the context of social network analysis entity association analysis, group identification, recommender system, prediction of links and attribute values.

The several machine learning algorithms were developed in every year. The quick enhancements

open up new implications. The algorithms are composed of following essential blocks:

Feature Selection: Many data sets contain several data fields. However, all fields are not equally important. Fields may be correlated. Thus the selection of the desired set of fields is essential to improve learning efficiency. Correlation matrix, information gain, gain ratio, principal component analysis etc. are being used for feature selection.

Choice of Algorithm : Several ML algorithms are available in the literature, but there is not a single algorithm that can fit all problems. Thus depending on the features of the dataset and domain, there is a need to select a specific ML algorithm. Decision trees, Naíve bayes, Support Vector Machines, k-Means Clustering are some regularly used ML algorithms. Training : ML improves a model by turning the model from the obtained error. The step is known as training.

Evaluation Criteria : To test a model, an ML algorithm requires a model evaluation criterion. Error rate, specificity-sensitivity, precision, recall, f1-score, robustness, etc are some commonly used methods. Based on the evaluation criterion, a stopping criterion must be defined by allowing some error. Otherwise, the training might take longer time than expected.

Testing : Testing of learned model is essential to validate the model. Cross-validation, leave-one-out are used to test the model.

There are four types of machine learning: Supervised learning (alternatively known as inductive learning): Learn a function from the datasets where class labels are defined. The main task is to classify or building models for predicting a future event. Classification and regression are the two supervised learning methods used for categorical and numerical data respectively. Unsupervised learning: Training data does not contain label.

The main task is to group data is known as clustering. Semi-supervised learning: Few training data are labelled.

Reinforcement learning: Learns by performing some action and then observing the action. Optimization is one of the key task of modern machine learning. Optimization relates to the process of discovering one or more feasible solutions which maximize or minimize an objective function under the given conditions. In this thesis, we have extensively use optimization in the form of both single objective and multi-objective. Succeeding sections discussed them.

The Multi-objective or the multi-criterion optimization problem is defined as the problem of finding a vector=

~x = [x1, x2, . . . , xd] T

in the feasible d-dimensional solution space Ω defined by the m inequality constraints

gi(~x) 6= 0, i = 1, 2, . . . , m; the p equality constraints

hi(~x) = 0, i = 1, 2, . . . , p; and that optimize k-objective vector function ~f(~x) = [f1(~x), f2(~x), . . . , fk(~x)]T , where, ~x = [x1, x2, . . . , xd] T is the vector of decision variables.

The problem can be mathematically expressed as:

Optimize ~x ~f (~x) = [f1 (~x), f2 (~x), . . . , fk (~x)]T Subject to gp(~x) ≤ 0, ∀p = 1, 2, . . . , n hq(~x) = 0, ∀q = 1, 2, . . . , m

A solution x (1) is said to dominate another solution x (2) , if both conditions 1 and 2 are true:

1. The solution x (1) is no worse than x (2) in all objectives, or fj (x (1)) χ fj (x (2)) for all j = 1, 2, . . . , K.

2. The solution x (1) is strictly better than x (2) in at least one objective, or fj (x (1)) $\bar{}$ fj (x (2)) for at least one j = 1, 2, . . . , K. The operators χ and $\bar{}$ are defined as follows: χ = (≤, Minimization Objective ≥, Maximization Objective $\bar{}$ = (, Maximization Objective)

The current focus on the study of environmental data in order to forecast future trends has elevated it to a top research priority worldwide. The current state of the data, as well as its future condition, can be forecasted or predicted, depending on the type of data prediction and analytical tools used. For the examination of data from the past or present in order to forecast future patterns, several approaches relating to modelling, statistics, data mining, artificial intelligence (AI), and machine learning are utilized. Defining the job, gathering associated data from various sources, assessing the data, statistical analysis, data modelling, deployment of the acquired data using diverse methodologies, and lastly, model monitoring are all processes involved in such analysis and forecasts. This type of predictive analysis is commonly used in a variety of scenarios, including market sales forecasting, consumer demand forecasting, healthcare status forecasting, collection analysis, fraud detection, and so on. Among them, the analysis and prediction of social media data is seen as a critical area of application, particularly for forecasting the future condition of highly contagious disease propagation. In this scenario, analyzing to related data to predict its spread and containment patterns is critical for halting the global users.

Methods for forecast prediction- Numerous types of analyses and predictions using information acquired from various sources such as daily updated WebPages, Kaggle, Orange, and Weka can be observed in order to achieve this criteria. As a result,

several strategies and methodologies developed by different researchers for projecting the future effects of the social media are considered as competing, each with its own set of strengths and flaws. Advanced Al approaches such as machine learning and deep learning have also been employed to carry out such forecasting, each with its own approach. Different methodologies and techniques, such as the regression model, the autoregressive model, the classification model, and so on, are employed in machine learning.

The methodology described and outlined in this paper is unique in that it will consider a method that combines nonlinear transmission with social-spatial and temporal transmission, as well as month-by-month data prediction. The majority of previous data-driven approaches will be linear, and do not take into account temporal or time-based transmission mechanisms.

SUPERVISED MACHINE LEARNING METHODS

To forecast and interpret future data, various supervised machine learning models will be deployed to get assurance about which models will be better for our study. We will target mainly Ensemble learning, autoregressive models, and moving average regressive models are among the models which will get employed in this work.

Learning in a group: Ensemble learning is the creation of a set of multiple models, such as experts or classifiers, to tackle a certain intelligence problem. Ensemble learning is primarily used to improve prediction, classification, and the construction of improved approximations of functions that must be learned. The prediction performance will improve using this strategy, and negative conditions resulting from the usage of weak predictions are avoided. For decision-making, incremental learning, and error correction, this learning model will be utilized. In this learning approach, “boosting” will be used to raise the weightage for misclassified training data, allowing the poor classifier to be reinforced. When this boosting approach will be used, accuracy will improve.

Autoregressive model: An autoregressive model predicts data based on time and observations from prior activities. Previous actions and the statistical association between previous observations are utilized to forecast data in the future.

The moving average model: The moving average model is a typical model for predicting data that is based on the linear relationship between current and previous values.

Steps to create a hybrid model for supervised learning:

The numerous procedures the prediction will be represented using hybrid model Algorithm which will be used in this research. The numerous input dataset sources will initially be included in the model’s training and testing.

The number of citations has been plotted: The fitted line (dotted) indicates that ‘as article getting older, the number of citations also increase’. It supports the social theory ‘rich get richer’. The correlation of article’s age and the number of citations has been shown. According to the pie chart, they are moderately and positively correlated with a degree of 0.42. Therefore, it can be stated that all articles are not equally significant for the progress of research. Although, good articles got citations along with their age.

CONCLUSIONS

The study was investigate “A machine learning methods for forecast prediction on social media users.” The structure of published research articles through the use of social network users. Therefore, with a complete case study highlighting steps of data collection and preprocessing, we have made three type of networks. The paper analysed them and summarized various statistical properties like Machine Learning Approach, Vector Machine, and Social media User. The study also have identified experts in the different category i.e., onmachine learning algorithms or techniques as Support Vector machine, Maximum Entropy. Naive Bayes and Decision tree. It was observed most of the groups are specialists in a single category. It is observed that people tried to ties with the authors having good publication. As have verified - age of publication and number of citation are positively correlated. Initially, it was thought that the subgraphs for a category must have the dense network, as compared to intra-connection. However, it is seen that the machine learning methods or pattern does not create such group of social media users.

REFERENCE

1. Bader, D.A., Kintali, S., Madduri, K. and Mihail, M., 2007, December. Approximating betweenness centrality. In International Workshop on Algorithms and Models for the Web-Graph (pp. 124-137). Springer Berlin Heidelberg.

2. Brandes, U., 2001. A faster algorithm for betweenness centrality*. Journal of mathematical sociology 25 (2), 163–177.

3. Dandekar, R., & Barbastathis, G. (2020). Quantifying the effect of quarantine control in cybercrime spread using machine learning, 34 (67), pp.43-53.

4. Freeman, L. C., 1977. A set of measures of centrality based on betweenness. Sociometry, 35–41.

5. Lee, M.J., Lee, J., Park, J.Y., Choi, R.H. and Chung, C.W., 2012, April. Qube: a quick algorithm for updating betweenness centrality. In Proceedings of the 21st

international conference on World Wide Web (pp. 351-360).

Corresponding Author

Renu Kumari*

College of Commerce Arts and Science, Patliputra University, Patna, 800020 Bihar, India