Journal of Advances in Mathematical & Computational Sciences
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA
© Creative Research Publishers
Available online at https://www.isteams.net/ mathematics-computationaljournal.info
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

Machine Learning & Associated Algorithms – A Review

[1]Sarumi, Jerry. Abayomi[1](PhD) & Oduroye, Ayorinde P[2] (PhD)
Department of Computer Science
Lagos State University of Science & Technology
Ikorodu, Lagos State, Nigeria, and Caleb University, Imota, Lagos State, Nigeria
E-mail: [1]jerrytechnologies@yahoo.co.uk
Phone: [1]+2348023408122

ABSTRACT

Machine learning and associated algorithms occupies a pride of place in the execution of automation in the field of computing and its application to addressing contemporary and human-centred problems such as predictions, evaluations, deductions, analytics and analysis. This paper presents types of data and machine learning algorithms in a broader sense. We briefly discuss and explain different machine learning algorithms and real-world application areas based on machine learning. We highlight several research issues and potential future directions

Keywords: Machine Learning, Algorithms, Applications, Predictions, Deductions, Analytics

Allenotor, D. & Oyemade, D. A. (2022): Machine Learning & Associated Algorithms – A Review. Journal of Advances in Mathematical & Computational Science. Vol. 10, No. 3. Pp 1-14
Available online at www.isteams.net/mathematics-computationaljournal.

1. MACHINE LEARNING

Machine learning and AI are the important concepts in the current scenario and machines will replace much of the human work. For example, in the future, bots will replace most of the humans in the armed forces of a country. Restaurants can replace the waiters with AI bots. Bots in restaurants are available in a few restaurants now. There are machine-learning approaches that can teach the bots to understand the environment and act accordingly. Classification, clustering, regression, and deep learning are some of the models in machine learning. As shown in Figure 4.1 the machine learning algorithms can be divided into four types, namely, supervised learning, unsupervised learning, semi supervised learning, and
Re-enforcement learning.

1.1 Machine Learning Techniques

Techniques of machine learning are fundamentally algorithms and this can be work on data and retrieve insights from it, this can include discovering, predicting, or forecasting patterns and trends. The concept is to develop a model utilizing a data combination and algorithms, which used to work on new, derive data before unseen data actionable insights. Every techniques relies on what type of data and the objective of issues they are trying to sort out. People generally get convinced to learn an algorithms couple and after that attempt to apply them to each issue (Nadikattu, 2017). An essential point to remember is that there is no exist any universal machine-learning algorithm that fixes complete issues. The main inputs for the machine-learning algorithm are aspects that retrieved from the data utilizing a procedure known as the aspects extractions, this is, generally coupled with the other process named aspect engineering or developing new features from the existing features.
Every feature can be explained as an aspect of the data set, like their locations, age, number of shares posts, and many more, if dealing with the data related to the social media profiles of the users. Techniques of machine learning can be categorized into two main types called supervised and unsupervised learning.

Supervised Learning

Supervised learning technique is a machine learning family subset algorithm which is mainly utilized in the forecast and predictive modeling. A predictive model is fundamentally a model that can be developed utilizing a supervised learning algorithm on aspects or features from the training data which is available data utilized to train or develop the model like which can forecast utilizing this model on newer, before unseen points of data (Xue et al., 2017). Algorithms of supervised learning attempt to model dependencies and relationships between target prediction result and input aspects like that they can forecast result value of the new data based on these relationships that learned from data set utilized during the model building or training. Supervised learning techniques categorized mainly in two types

a) Classification: These types of algorithms develop predictive models from the training data where the feedback variable to be forecasted is unconditional. These all forecasted models utilize the aspects learned from the training data on new; unseen data, which predict their category labels. The result classes relate to the discrete classifications. The algorithm involves support vector machines, decision trees, random forests, and much more.
b) Regression: This type algorithm is utilized to develop a predictive model on data like which the feedback variable to be forecasted is numerical. The algorithms develop a model based on the input attributes and output feedback values for the training data and the model utilized to forecast values for the new data. The result value in this case is regular numeric values and not discrete classifications. Regressions type algorithm involve the linear regression, multiple regression, lasso regression, and ridge regression among the many other types.

Unsupervised Learning

These techniques are the subset of machine teaching algorithms family that is mainly utilized in the dimension reduction, descriptive modeling, and pattern detection. The descriptive model is fundamentally a model built by an unsupervised machine learning algorithm and aspects from the input data the same to supervised learning procedures. But the result feedback variables do not exist right now in this case. These algorithms attempt to utilize techniques on the input data to mine for the rule and regulations, summarize and group data points and detect patterns that help in deriving significant insights and explain the data better to users. unsupervised learning techniques are categorizes-

i. Clustering
ii. Association rule mining
iii. Dimensionality reduction

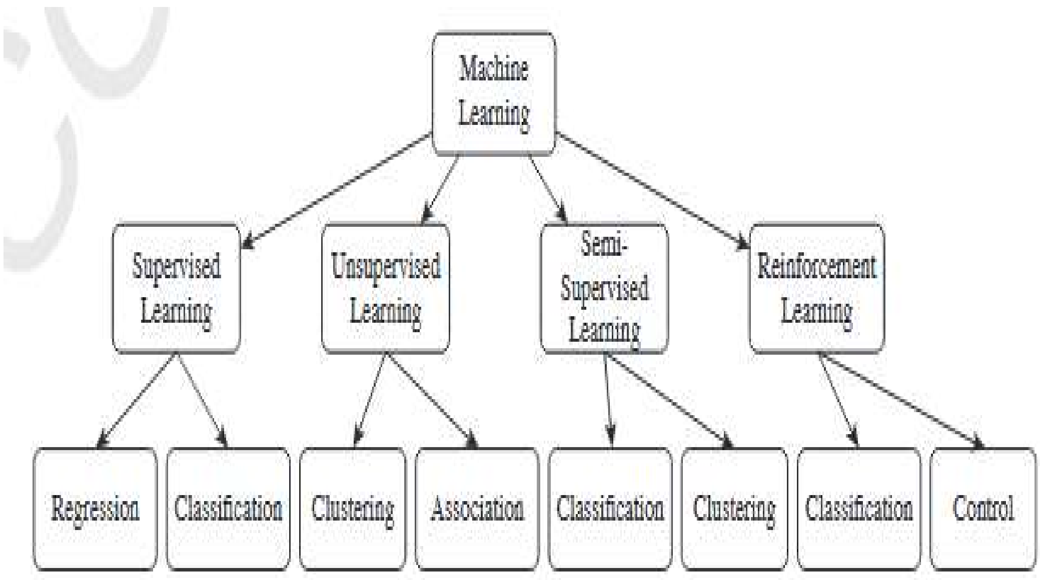

Figure 1: Classification of machine learning algorithms

2. MACHINE LEARNING TASK AND ALGORITHMS

In this section, we discuss various machine learning algorithms that include classification analysis, regression analysis, data clustering, association rule learning, feature engineering for dimensional reduction, as well as deep learning methods. A general structure of a machine learning-based predictive model has been shown in fig 2 where model is trained from historical data in phase 1 and the outcome is generated in phase 2 for the new test data

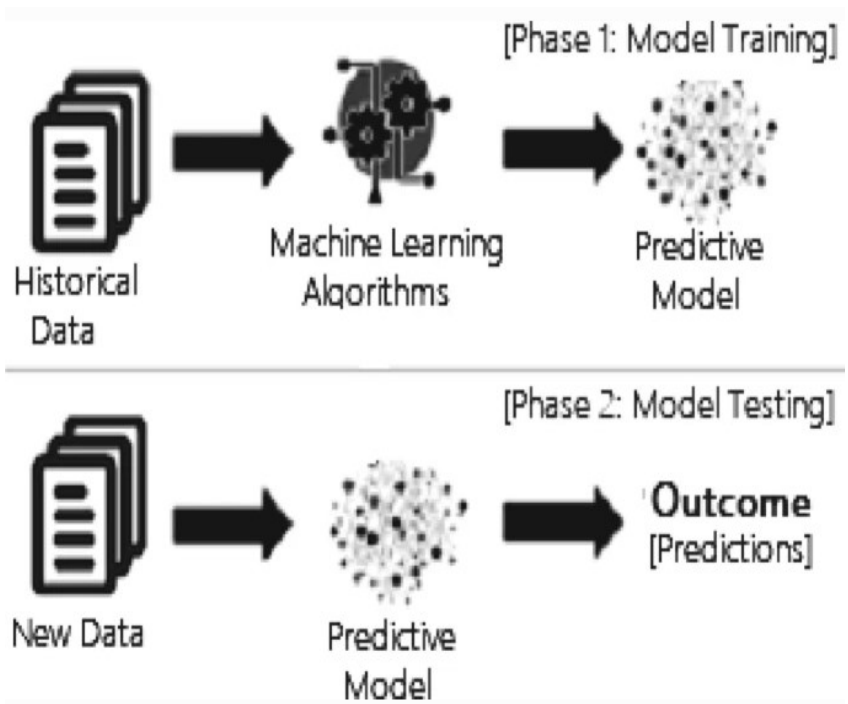

Figure 2: General Structure Of A Machine Learning Based Predictive Model Considering Both The Training And Testing Phase

Classification and Regression

In classification the dotted line represents a linear boundary that separates the two classes. In regression the dotted line models the linear relationship between two variables as shown in figure 2.6

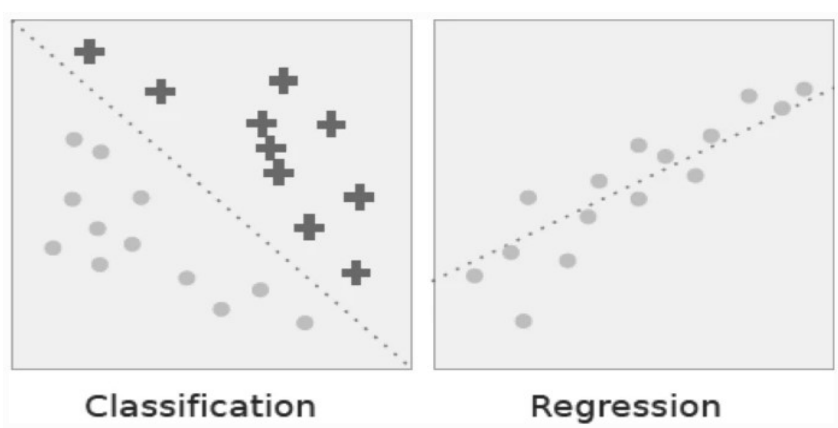

Figure 3: classification vs regression

Deep Learning and Social Networks

Deep learning was first used in social networks by Perozzi et al. (2014). In this work, the authors used deep learning to represent social graphs with a latent representation in continuous vector space. This allows other well-known statistical and machine learning models to be used with social network data easily. To learn the social representation, they used a stream of short random walks. In 2015, Nikfarjam et al. (2015) used deep learning tech

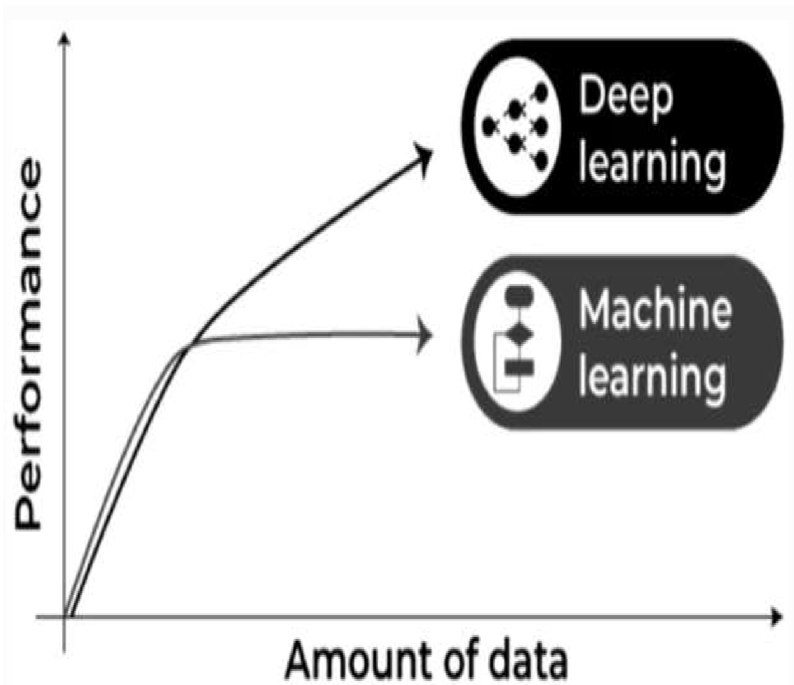

Figure 4: Machine learning and deep learning performance in general with the amount of data

Social Media Algorithms

Social media algorithms are a way of sorting posts in a users' feed based on relevancy instead of publish time. Social networks prioritize which content a user sees in their feed first by the likelihood that they'll actually want to see it.

Before the switch to algorithms, most social media feeds displayed posts in reverse chronological order. In short, the newest posts from accounts a user followed showed up first. This is still an option on Twitter to set your feed to chronological order

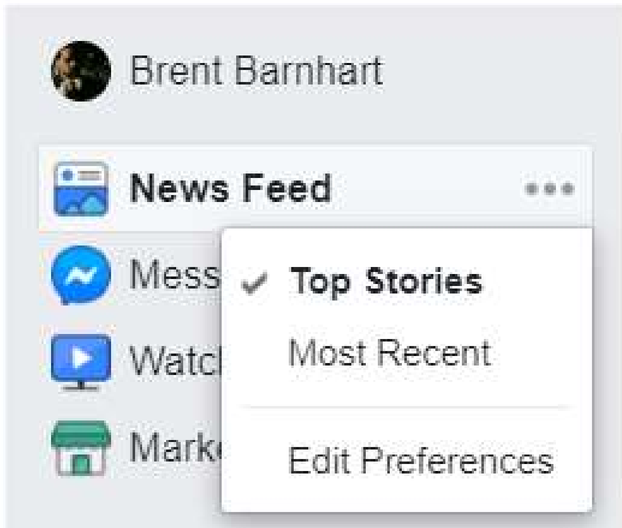

Figure 5: News Feed

By default, social media algorithms take the reins of determining which content to deliver to you based on your behavior. For example, Facebook or Twitter might put posts from your closest friends and family front-and-center in your feed because those are the accounts you interact with most often. Chances are you've been recommended videos to watch on YouTube, right? This is again based on your individual behavior, digging into what you've watched in the past and what users like yourself are watching. Elements such as categories, #tags and keywords also factor into recommended content on any given network.

3. WHY SOCIAL MEDIA ALGORITHMS EXIST

There is a *ton* of content floating around in the social space. Like, thousands of posts, photos and videos published per minute. Without social media algorithms, sifting through all of this content on an account-by-account basis would be impossible. Especially for users following hundreds or thousands of accounts on a network, so algorithms do the legwork of delivering what *you* want and weeding out content that's deemed irrelevant or low-quality. In theory, that is there's also the belief that social media algorithms exist to push brands to pay a premium for social ads.

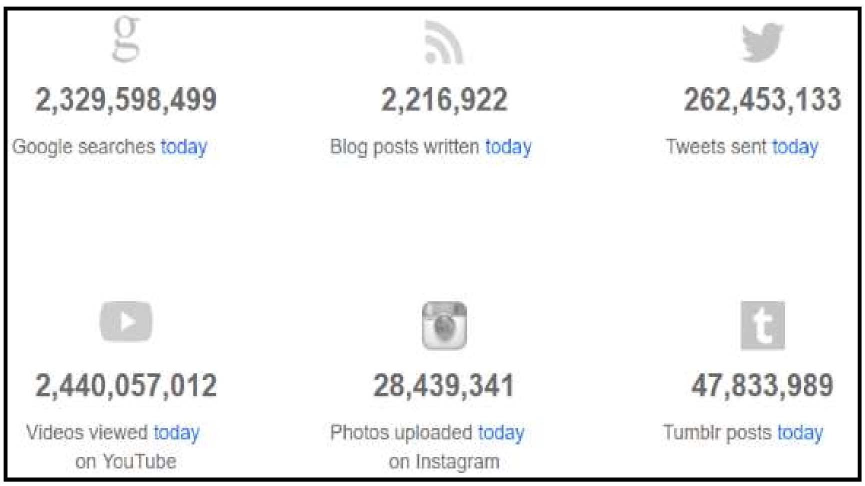

Figure 6: Social Media Post

The belief is that if brands can't reach their audience organically, they'll turn to ads instead. Obviously, this means more money for social networks. This point-of-view might seem cynical or even paranoid, but social marketers know that changes in how social media algorithms prioritize paid and organic content can have a huge impact. Regardless of why social media algorithms exist, the fact remains that they aren't going anywhere. For brands, this means learning what algorithms "want" and likewise what might cause content to be viewed as low-quality or irrelevant to their audience

How social media algorithms work

For starters, social media algorithms treat engagement (think: likes, comments and shares) as a sort of snowball effect. That is, the more engagement a piece of content gets, the more likely it is to be rewarded by the algorithm. Perhaps one of the easiest ways to encourage engagement is by asking questions of your followers. Serving as a sort of call-to-action, question-based posts are an easy way to encourage interactions and connect with your audience at the same time.

4. SENTIMENT ANALYSIS

In the age of digitalization, a huge amount of sentiments are expressed daily on university related topics using social media platforms. Particularly, posted statements from students and teachers can provide a potential source for evaluating universities. Twitter as one of the most popular microblogging platforms is a rich data resource for opinion mining. Stimulated by this fact, ways to analyze Twitter for information in the context of academics are sought.

Sentiment analysis is done using natural language processing and information extraction with the goal of obtaining the writer's feeling as positive, negative or neutral (Subhabrata, 2017). Sentiment analysis is often used as component in opinion mining when the goal is to is to analyze sentiment and attitudes (Bing, 2017). There are a number of various methods that can be used to classify sentiment of a text. In this project the Dictionary-based approach is used for sentiment analysis.

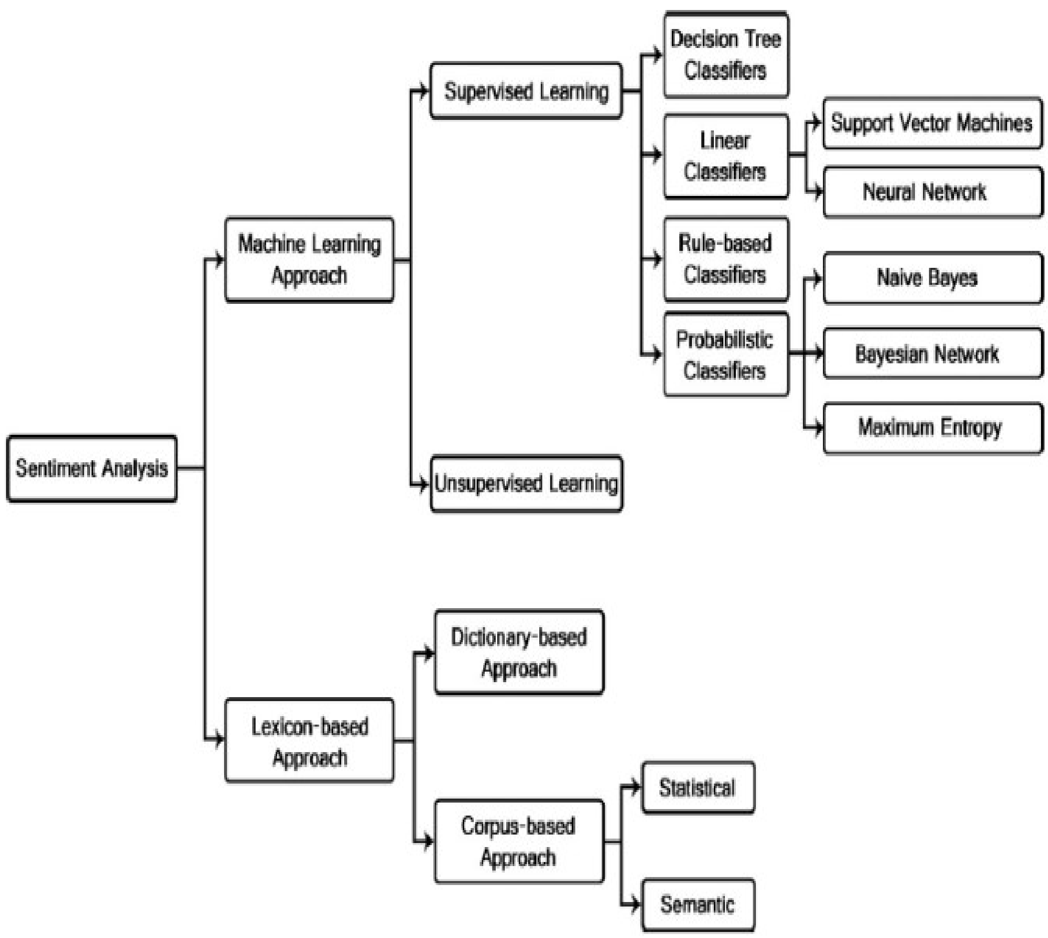

Figure 7: Sentiment Analysis Methods (Bing, 2017)

Sentiment Analysis for Social Media

Social Sentiment analysis is the use of natural language processing (NLP) to analyze social conversations online and determine deeper context as they apply to a topic, brand or theme. Our net sentiment score and brand passion index show how users feel about your brand and compares across your competitors (Nawaz et al., 2019). Emotions – big bold, volatile, passionate emotions – are the driving force behind everything social users share. If you don't have a way to analyze all that shared sentiment, you can't possibly know what matters most to the people you want to love your brand. To make the most of Social Media Sentiment Analysis, you've got to monitor conversations to learn two key things:

- Whether consumers' emotions are positive or negative (Net Sentiment).
- How strong those emotions are (Passion Intensity).

Sentiment analysis for social media marketing

Sentiment analysis, also called opinion mining or emotion AI, is the practice of judging the opinion of text data. The process uses both natural language processing (NLP) and machine learning to pair social media data with predefined sentiment labels such as positive, negative or neutral. Then, the machine can develop agents that learn to understand the underlying sentiments in new messages. Businesses can apply sentiment analysis in social media and customer support to collect feedback on a new product, service or design. Similarly, businesses can apply sentiment analysis to discover how people feel about their competitors or trending industry topics.

Sentiments Analysis for Social Media Academics.

Analysis of the sentiments, also named mining of opinion or the emotion AI, is judge opinion of the text. The procedures utilize both natural language processing (NLP) and ML (machine learning) to the pair of social media data with predefined labels like negative, positive, or neutral. Next, the machine can make agents who learn to understand the sentiments that underlying new messages. Education can be applied sentiment examination on social media and customer support to gather responses on the new design or product. Comparably, education can apply the sentiments analysis to find out how people feel their institutions or topics about the trending industry (Xue et al., 2017).

5. DATA VISUALIZATION TOOLS

Data visualization tools provide business intelligence (BI) capabilities and allow different types of users to gain insights from the big data. The users can perform explanatory analysis through interactive user interfaces available on the majority devices, with a recent focus on mobile devices (smartphones and tablets). The data visualization tools help the users identify patterns, trens and relationships in the data which were previously latent, fast ad hoc visualization on the data sets frameworks.

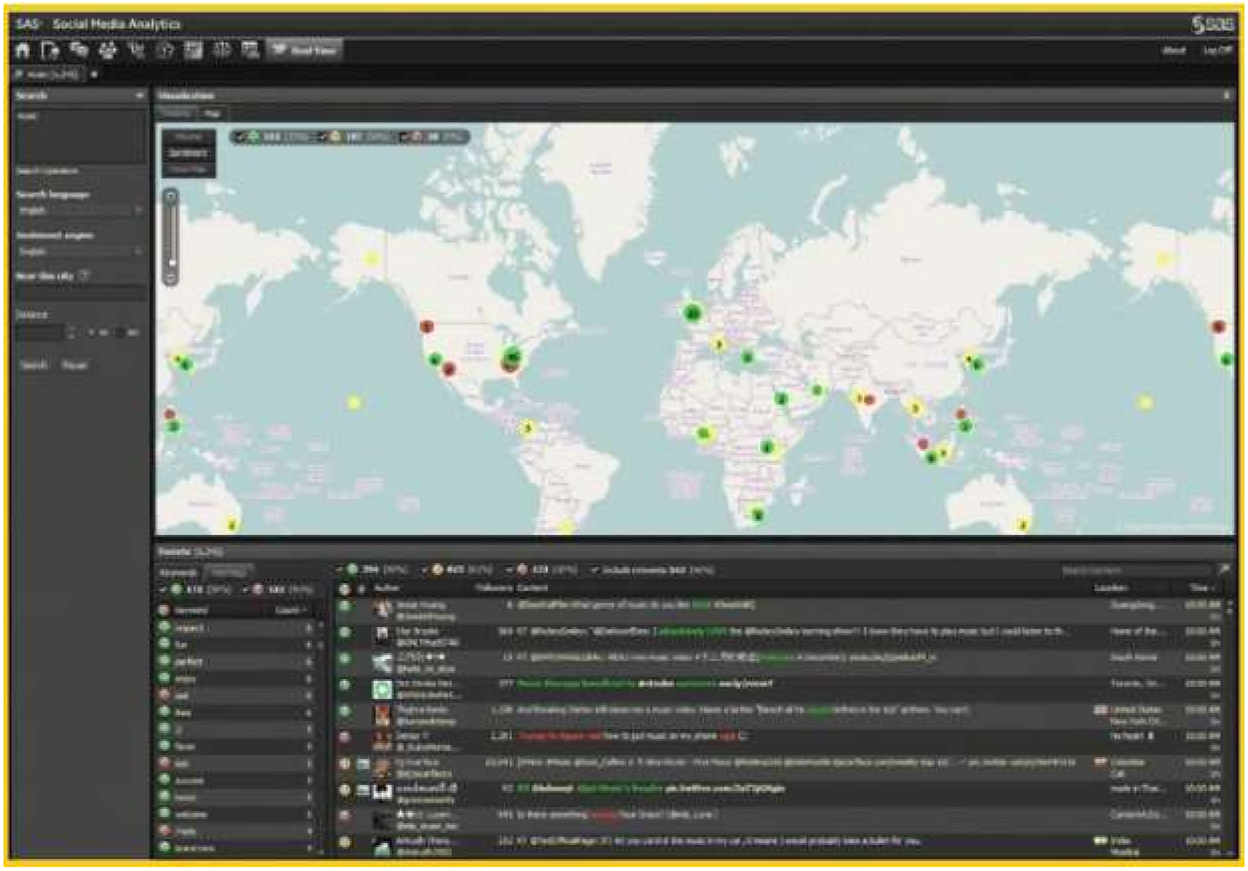

Figure 8: SAS Visualization of Real-Time Tracking via Twitter

6. NATURAL LANGUAGE PROCESSING

The amount of content generated by the users of social media is exponentially increasing. The text data cannot be processed by a machine efficiently like with other formats of data. A machine needs to understand human slang and language to analyze the text content. Natural language processing (NLP) helps machines understand human slang and language in the text content generated on social media. The flow of content from social media to a big data storage system and the analysis by ML and NLP are illustrated in Figure 9.

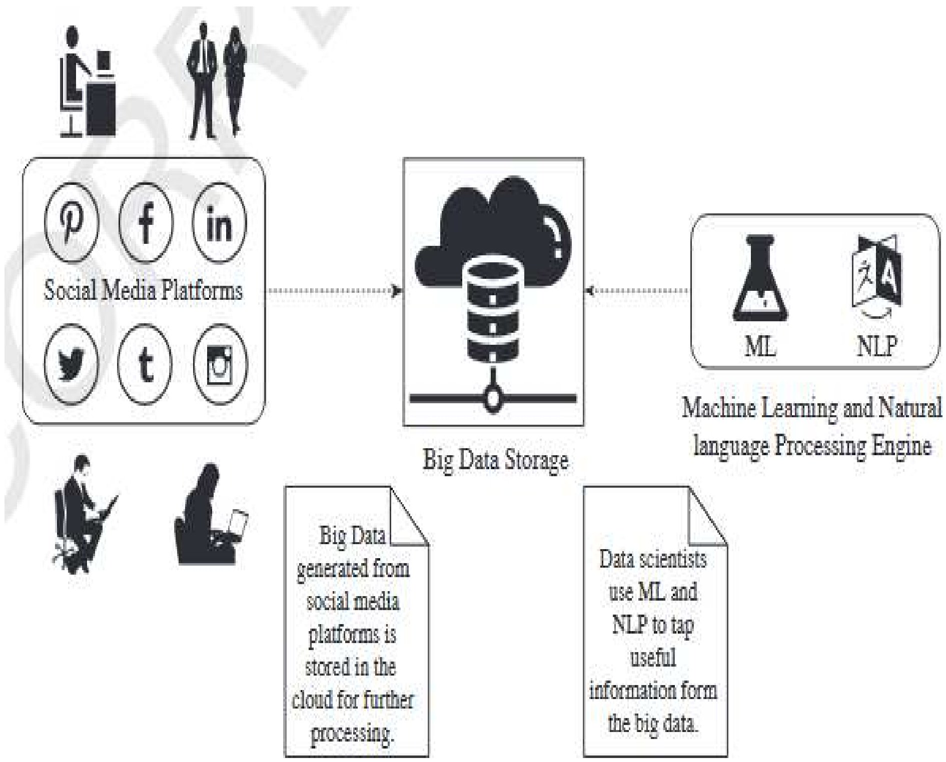

Figure 9: Workflow of Big Data, Machine Learning and Social Media

7. CONCLUSION

Machine learning algorithms can perform tasks without being explicitly programmed to do so. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step. The discipline of machine learning employs various approaches to teach computers to accomplish tasks where no fully satisfactory algorithm is available. In cases where vast numbers of potential answers exist, one approach is to label some of the correct answers as valid. This can then be used as training data for the computer to improve the algorithm(s) it uses to determine correct answers.

BIBLIOGRAPHY

1. Reza Zafarani, Mohammad Ali Abbasi, Huan Liu (2019). Social Media Mining Cambridge University
2. Buchem, I., Ebner, M., Schön, S., Appelt, R., & Kaiser, S. (2018). Blogging und Microblogging. Lehrbuch für Lernen und Lehren mit Technologien: 2. Auflage (2013), 239.
3. Gapski, H., & Gräßer, L. (2017). Medienkompetenz im web 2.0–lebensqualität als zielperspektive. Praxis Web, 2, 11-34.
4. Hisserich, J., & Primsch, J. (2020). Wissensmanagement in 140 Zeichen: Twitter in der Hochschullehre. Community of Knowledge.
5. Maynard, D., & Funk, A. (2016). Automatic detection of political opinions in tweets. In Extended Semantic Web Conference (pp. 88-99). Springer, Berlin, Heidelberg.
6. Go, A., Bhayani, R., & Huang, L. (2019). Twitter sentiment classification using distant supervision. CS224N project report, Stanford, 1(12), 2009.
7. Atefeh, F., & Khreich, W. (2015). A survey of techniques for event detection in twitter. Computational Intelligence, 31(1), 132-164.
8. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. J. (2016). Sentiment analysis of twitter data. In Proceedings of the workshop on language in social media (LSM 2011) (pp. 30-38).
9. Kouloumpis, E., Wilson, T., & Moore, J. (2016). Twitter sentiment analysis: The good the bad and the omg!. In Fifth International AAAI conference on weblogs and social media.
10. Pak, A., & Paroubek, P. (2020). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).
11. Davidov, D., Tsur, O., & Rappoport, A. (2020). Enhanced sentiment learning using twitter hashtags and smileys. In Coling 2010: Posters (pp. 241-249).
12. Kaplan, A. M., & Haenlein, M. (2020). Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, 53(1), 59-68.
13. Subhabrata Mukherjee. (2017). Sentiment analysis. Indian Institute of Technology, Bombay Department of Computer Science and Engineering.
14. Hagel, J. I. I. I., Brown, J. S., & Davison, L. (2020). The best way to measure company performance. Harvard Business Review, 4(6), 56-60.
15. Needles, B. E., Powers, M., & Crosson, S. V. (2018). Principles of accounting. Cengage Learning.
16. Granger, C. W. (2019). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society, 424-438.
17. Michael T. Lash and Kang Zhao (2016). Early Predictions of Movie Success: the Who, What, and When of Profitability Artificial Intelligence (cs.AI); Social and Information Networks (cs.SI).
18. Sahar Nassirpour, Parnian Zargham, Reza Nasiri Mahalati (2017). Electronic Devices Sales Prediction Using Social Media Sentiment Analysis Project report Stanford university.
19. Richard Socher Jeffrey Pennington, Eric H. Huang Andrew, Y. Ng Christopher, D. Manning (2016). Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions Proceedings of the Conference on Empirical Methods in Natural Language Processing Pages 151- 161

20. Tristan Fletcher, Fabian Redpath and Joe DAlessandro (2019) Machine Learning in FX Carry Basket Prediction Proceedings of the International Conference of Financial Engineering, vol. 2, page 1371-1375.
21. Tim Loughran and Bill McDonald (2016) When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks The Journal of Finance, Vol. LXVI, NO. 1
22. Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining.In LREc (Vol. 10, No. 2010, pp. 1320-1326).
23. Douglas M. McNair, Maurice Lorr, and Leo F. Droppleman (2021). Manual for the Profile of Mood States San Diego, CA: Educational and Industrial Testing Service
24. Carol McDonald. (2015). Parallel and Iterative Processing for Machine Learning Recommendations with Spark https://www.mapr.com/blog/ parallel-and-iterative-processing-machine-learning-recommendations-spark
25. Karina Gibert, Miquel Snchez-Marr, Vctor Codina. (2020). Principles of Accounting. International Environmental Modelling and Software Society (iEMSs).
26. Rokach, L., & Maimon, O. Z. (2017). Data mining with decision trees: theory and applications (Vol. 69). World scientific.
27. Freund, Y., & Schapire, R. E. (2016). Experiments with a new boosting algorithm. In icml (Vol. 96, pp. 148-156).
28. Sylvain Arlot, (2019). A survey of cross-validation procedures for model selection. Journal of Machine Learning Research , p1089-1105.
29. Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer. (2017). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, page 321357.
30. Speriosu, M., Sudan, N., Upadhyay, S., & Baldridge, J. (2016). Twitter polarity classification with label propagation over lexical links and the follower graph. In Proceedings of the First workshop on Unsupervised Learning in NLP (pp. 53-63).
31. Medhat, W., Hassan, A., & Korashy, H. (2019). Sentiment analysis algorithms and applications:
32. A survey. Ain Shams engineering journal, 5(4), 1093-1113.
33. Davidov, D., Tsur, O., & Rappoport, A. (2020). Enhanced sentiment learning using twitter hashtags and smileys. In Coling 2010: Posters (pp. 241-249).
34. Read, J. (2015). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In Proceedings of the ACL student research workshop (pp. 43-48).
35. Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2016). Combining lexicon-based and learning-based methods for Twitter sentiment analysis. HP Laboratories, Technical Report HPL-2011, 89.
36. Onan, A., Korukoğlu, S., & Bulut, H. (2016). A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. Expert Systems with Applications, 62, 1-16.
37. Yoon, V. Y., Hostler, R. E., Guo, Z., & Guimaraes, T. (2018). Assessing the moderating effect of consumer product knowledge and online shopping experience on using recommendation agents for customer loyalty. Decision Support Systems, 55(4), 883-893.
38. Lin, J., & Kolcz, A. (2017). Large-scale machine learning at twitter. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (pp. 793-804).

39. Rodriguez-Penagos, C., Atserias, J., Codina-Filba, J., García-Narbona, D., Grivolla, J., Lambert, P., & Saurí, R. (2018). Fbm: Combining lexicon-based ml and heuristics for social media polarities. In Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2018) (pp. 483-489).
40. Hassan, A., Abbasi, A., & Zeng, D. (2018). Twitter sentiment analysis: A bootstrap ensemble framework. In 2018 international conference on social computing (pp. 357-364). IEEE.
41. Clark, S., & Wicentwoski, R. (2018). Swatcs: Combining simple classifiers with estimated accuracy. In Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2018) (pp. 425-429).
42. Jianqiang, Z. (2016). Combing semantic and prior polarity features for boosting twitter sentiment analysis using ensemble learning. In 2016 IEEE first international conference on data science in cyberspace (DSC) (pp. 709-714). IEEE.
43. Pang, B., Lee, L., & Vaithyanathan, S. (2017). Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070.
44. Onan, A., Korukoğlu, S., & Bulut, H. (2016). A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. Expert Systems with Applications, 62, 1-16.
45. Prusa, J., Khoshgoftaar, T. M., & Dittman, D. J. (2015). Using ensemble learners to improve classifier performance on tweet sentiment data. In 2015 IEEE International Conference on Information Reuse and Integration (pp. 252-257). IEEE.
46. Coletta, L. F., da Silva, N. F., Hruschka, E. R., & Hruschka, E. R. (2019). Combining classification and clustering for tweet sentiment analysis. In 2014 Brazilian conference on intelligent systems (pp. 210-215). IEEE.