

Machine Learning as A Tool for Analysing Hotel Online Reviews

Pankaj Chaudhary
Research Scholar
ICFAI University, Dehradun

Dr. Anurag Aeron,
Associate Professor(CSE)
ICFAI University, Dehradun

Dr. Sandeep Vijay
Director, Shivalik College
of Engineering, Dehradun

Abstract:- During current days once somebody attempt to book a edifice, previous on-line reviews of the edifices play a serious role in the decisive parameters to select the budget for the client. Previous on-line reviews play the foremost vital motivation for the knowledge and hotel business growth. Owing to the high impact of the reviews on business, hotel owners invariably extremely involved and centered regarding the client feedback and past on-line reviews. however all reviews aren't true and trustworthy, someday few individuals might by choice generate the faux reviews to form some hotel popular. So it's essential to develop and propose the techniques for analysis of reviews. With the assistance of assorted machine learning techniques viz. supervised machine learning technique, Text mining, unsupervised machine learning technique, Semi-supervised learning, Reinforcement learning etc we have a tendency to detect the faux reviews. This paper provides some notions of identification of appropriate machine learning techniques in analysis of past on-line reviews of hotels, supported the observation it additionally counsel the best machine learning technique for particular hotel reviews groups.

Keywords: Machine learning technique, Text mining, supervised machine learning technique, Semi-supervised machine learning technique, Reinforcement machine learning technique, Hype, Quantification, collision, manipulation, machine learning, mining, deep learning etc

I. IMPORTANCE OF ONLINE REVIEWS

1.1 Faux reviews have serious consequences.

In case of hotels a lot of positive reviews earn a lot of reservations and business. It is tempting to request friends, family, and staff to give away positive reviews on-line for the hotels or perhaps to procure high marks on-line. However, apart from being unethical and deceptive, faux reviews will have serious consequences. Supposed we wish to travel abroad, faux reviews will virtually spoil our move expertise in a very new country .

1.2 Counterfeit reviews will backfire .

People can quickly be ready to conclude that a review was faked. If the expertise represented within the review is wildly completely different from the customer's expertise, they're going to like a shot suspect that they were deceived. programme Land represented a state of affairs wherever one company was commerce reviews. In alternative words, they offered to put in writing a decent review for a business if somebody would conform to write a decent review for them. Instead, individuals wrote 1-star reviews career them out for the phony observe. If a client suspects that any of your

positive reviews is a faux, then they're going to lose trust within the alternative positive reviews and can assume the negative reviews is the reality. you may hurt your name each on-line and offline.

1.3 Phony reviews are outlaw .

In spite to the intense dangerous strict actions unacceptable against those that write fraud evaluations. But bit by bit currently the situation is dynamic day by day. e.g. the owner of a business that sold-out faux favourable TripAdvisor reviews to hotels and restaurants in Italy has been confined for 9 months for Travel and Leisure, UN agency owns PromoSalento, was additionally awarded a Rs six,66,821 fine to hide for the damages done. The sentence was awarded by the court of Lecce. TripAdvisor was alerted regarding PromoSalento in 2015, once multiple hotels and restaurants forwarded them the company's e-mails providing faux reviews. The hotels had apparently received bulk e-mails from the corporate, wherever they'd offered nice reviews on TripAdvisor in exchange for cash. the corporate had additionally submitted regarding faux reviews on many hotels and restaurants, boosting their property ratings. when TripAdvisor found the offender, they demoted all the places UN agency had taken facilitate from PromoSalento. One automotive concern cluster was punished \$3.6 million by the FTC for reasons together with faux reviews? meantime, one business owner in USA was recently given a nine-month jail sentence for writing many faux reviews. Soliciting faux reviews could be a deceptive observes that interferes with a human ability to form a shopping for call. Faux reviews as fallacious info about your business, they become a lot of less appealing. Additionally, if to contemplate incentivizing customers for reviews (for example, by providing a reduction in exchange for a review), this info should be disclosed within the review .

1.4 Counterfeit reviews can get caught .

Review websites ought to defend their own name to still be trustworthy by shoppers. that is why Yelp and TripAdvisor have subtle code and algorithms to flag suspicious reviews. They then add warning messages to business pages to tell shoppers that they believe suspicious review activity has taken place. In fact, Yelp goes as way on run sting operations to uncover businesses that place ads to procure faux reviews. If the business is found to be soliciting paid reviews, it'll get this pop-up on the business listing: TripAdvisor and Yelp are within the business of serving to the buyer 1st, thus they're going to err on the facet of caution in terms of review

legitimacy. a technique that TripAdvisor identifies faux reviews is by pursuit information processing addresses of reviewers. which means that if a edifice receives multiple reviews from the information processing address, it's going to be assumed that staff at the property are writing reviews.

II. SOME COMMON TECHNIQUES

To prevent being de jure flagged as faking reviews, request reviews from guests when they properly stay and when they have left the property.

2.1 Superlatives

Fake reviewers are the professionals or members of the family of the property owner. Either way, one Brobdingnagian giveaway is that they tend to overuse superlatives and exaggeration. They use "awesome", "really" and "very" words. And look ahead to exclamation marks!! They'll cowl details that your regular guest wouldn't even notice and err on the facet of sounding sort of a gushing folder.

2.2 Compare sites

TripAdvisor permits anyone to post a review, whereas Expedia allows solely people that have set-aside through their web site to post. It stands to reason that those that have set-aside and paid can be additional reliable. A University of Southern California, Yale and Dartmouth study found that smaller, freelance hotels had additional five-star reviews on TripAdvisor than on Expedia. one set-aside an area through VRBO and wasn't allowed to transfer negative review. when escalating it to management (and receiving 2 phone calls from military headquarters in Dallas) one used to be told "this is associate degree advertising site".

2.3 Fewer real expertise reviews

Fake reviewers pay very little time writing regarding their impressions of the place (because they're in all probability sitting on the other facet of the globe and paid to blitz an internet site with a certain range of reviews by lunchtime). A Cornell University study assigned students the task of writing pretend reviews for four hundred Chicago hotels. Twelve per cent of them spent one minute writing each. They have a tendency to feature additional "I" and "me" to hide their circumstances instead of any specific detail regarding the hotel, the room, the facilities, the amenities

2.4 Timing

Some corporations rent writers to write pretend reviews. Check the time stamps on once the majority were denote on sites like Tripadvisor or AirBnB or VRBO as several contractors get 48 hours to show some hundred around. If you're suspicious, you can additionally click on the reviewer profile to measure however honest they seem on a variety of alternative properties. This story within the Daily Mail shows however associate degree shrewd Indian hotel owner checked the author profile and located he'd denote four unhealthy reviews of Indian restaurants on constant day. Tripadvisor eventually force the reviews.

2.5 Scan all of them

Scan and also the manager's response The biggest tip is to scan through countless reviews for the property you're fascinated by. Cull the over-enthusiastic ones and the miserable ones, presumptuous they need been written by the hotel's workers and their humorless competitors, and what you're left with can doubtless be the honest average. And if you see this: "TripAdvisor has affordable cause to believe that either this property or people related to the property may have tried to control our quality index by interfering with the unbiased nature of our reviews.

III CITIES HAVING THE VERY BEST SPURIOUS BUILDING REVIEWS

More than common fraction of on-line building reviews area unit, new analysis from fraud detection company Fakespot suggests that the disturbing study of faux building reviews additionally shows that you are additional doubtless to be fooled in some yankee cities than others. Fakespot, that analyzes and detects suspicious on-line reviews, says 12 months of the ratings on TripAdvisor area unit "fake and unreliable." to induce to the primary page these businesses can use pretend reviews to extend their search ranking." TripAdvisor laid-off the findings as "inaccurate and deceptive." currently, for the primary time, Fakespot has free an inventory of cities with the foremost pretend building reviews. It offers a guide for summer travellers WHO are attempting to avoid falling prey to a deceitful review.

3.1 Houston (38.4%)

H-Town topnotch the list of cities with the foremost pretend building reviews. "The newer cities like Houston don't have as several established businesses because the previous ones do," says kalif. "The landscape of the welcome trade is continually experiencing growth and volatility." That volatility, he says, makes it a main target for a pretend review. Example: AN extraordinary review of the building Alessandra, that declared, "I idolized each moment of being there over the weekend! the placement, the staff, the rooms, everything was amazing! The bell hop men area unit THE BEST!!!!!!!" The building Alessandra says the review is legitimate.

3.2 Chicago (35.9%)

While Chicago's pretend building review numbers area unit high, they're common fraction of a share below the national average. meaning outside of the massive cities, wherever new hotels area unit being designed, there is a higher probability of finding a pretend building review. Still, the Windy City's numbers area unit high -- too high for a longtime geographical area within the us,. The reviews for the grounds by Marriott Chicago Downtown/Magnificent Mile received a "D," with solely 0.5 the reviews listed as reliable

3.3 Miami (34.8%)

Fake building reviews abound here take into account this rating for the National building in metropolis, that Fakespot flagged as spurious. Superlatives like "amazing," "beautiful," and "wonderful" mean that a full hour of the reviews area unit

pretend. "Miami could be a terribly competitive building market, with a great deal of properties vying for ad," says kalif.

3.4. Dallas (34%)

Khalifah says city additionally features a heap of faux building reviews as a result of it's several new properties. Here's one for the furniture city that raised Fakespot's red flag: "I was on property last week for a fast business trip and was astonished by the standard of the merchandise and also the worth provided." Fakespot gave the Sheraton's reviews AN overall grade of "C."

3.5. Boston (30.5%)

Boston has thirty point five % pretend reviews. Here pretend reviews area unit longer long. "This low key building has everything you would like. Workability is incredibly high with masses to try and do, see, and dine in shut proximity. building workers was friendly and useful with all our desires...even at odd hours."

3.6. Loss Angles (30%)

There area unit approx half reviews that area unit pretend. e.g. this review of the London in West Hollywood, that raved regarding its "amazing" client service. Fakespot says it's most likely a pretend

3.7. New York (29.5%)

New York has approx twenty nine point five % phony reviews as per Fakespot. "The well-known hotels here do not vie exploitation pretend reviews. "They're likely to rank extremely thanks to their name." however AN occasional pretend slips in. as an example, here's a review of the Artezen building that plumbed many alarms: "The lobby smells delicious that could be a good way to be greeted, and also the workers go the additional mile to produce true welcome, one thing that's usually lacking during this town."

3.8. San francisco (28.5%)

San francisco has approx twenty eight point five % phony reviews as per Fakespot. Consider , some reviews that raves that it is a "fabulous" building. "The workers create the place exceptional, terribly hospitable, super useful and invariably friendly,". Fakespot disagrees, giving it a "C" for credibleness.

IV SUPERVISED LEARNING

Supervised learning refers to the task of inferring a performance from tagged set of knowledge. Equations square measure wont to fitting tagged coaching data; and also the purpose is to seek out the foremost best model parameters to predict unknown labels on alternative objects. Just in case label could be a real, then this task is termed as regression. If the label is within the sort of the restricted range of un-ordered values, then this can be referred to as classification. Numerous supervised learning techniques square measure as follows:

4.1. Decision Trees

A decision tree could be a decision support tool that uses a

tree-like graph or model of choices and their doable consequences, as well as chance-event outcomes, resource prices, and utility. From a business call purpose of read, a choice tree is that the minimum range of yes/no queries that one needs to raise, to assess the likelihood of constructing an accurate call, most of the time. As a way, it permits you to approach the matter in an exceedingly structured and systematic thanks to make a logical conclusion.

Advantages of Decision Tree Machine learning algorithm

These machine learning algorithms facilitate to build choices in below uncertainty and assist to improve communication, as they gift a visible illustration of a call state of affairs. Decision tree machine learning algorithms facilitate a knowledge human capture the concept that if a unique call was taken, then however the operational nature of a state of affairs or model would have modified intensely. Decision tree algorithms facilitate build best choices by permitting knowledge human to traverse through forward and backward calculation ways. E.g.

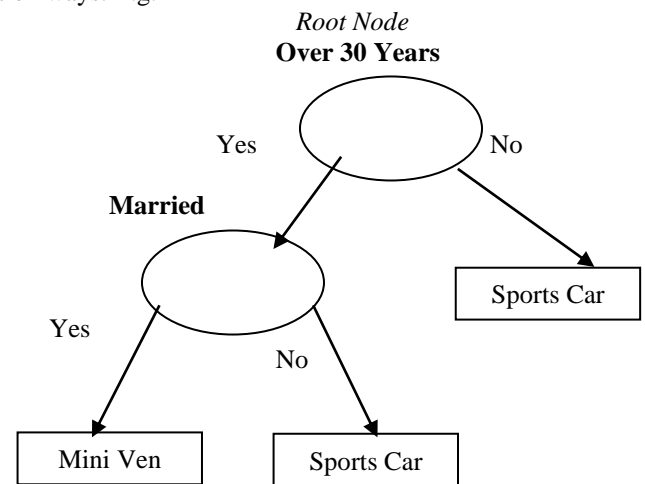


Fig 4.1: Decision Tree

4.2. Naive Bayes Classification: Naive Bayes classifiers technique is based on simple Bayes' theorem assuming strong independence assumptions between the features. P(A|B) is the probability of happening of A when B has already occurred, We may use Bayes's Theorem as follows: probability of hypothesis H to be true can be given as below.

$$P(H) = P\left(\frac{H}{A}\right).P(A) + P\left(\frac{H}{B}\right).P(B)$$

where:

- P(A|B) = Posterior probability. The probability of hypothesis A being true, given the data B, where P(A|B)= P(A1| B) P(A2| B)...P(An| B) P(B)
- P(A|B) = Likelihood. The probability of data B given that the hypothesis A was true.
- P(A) = Class prior probability. The probability of hypothesis A being true (irrespective of the data)

- $P(B)$ = Predictor prior probability. Probability of the data (irrespective of the hypothesis)

This algorithmic program is termed 'naive' as a result of it assumes that everyone variable square measure freelance of every different, that could be a naive assumption to form in real-world examples. a number of globe examples are:

1. To mark Associate in nursing email as spam or not spam
2. Classify a article regarding technology, politics, or sports
3. Check a bit of text expressing positive emotions, or negative emotions.
4. Used for face recognition software package.

Advantages of the Naïve Thomas Bayes Classifier Machine Learning algorithmic

Naïve Thomas Bayes Classifier algorithmic program performs well once the input variables square measure categorical. A Naïve Thomas Bayes classifier converges quicker, requiring comparatively very little coaching knowledge than different discriminative models like provision regression, once the Naïve Thomas Bayes conditional independence assumption holds. With Naïve Thomas Bayes Classifier algorithmic program, it's easier to predict category of the check knowledge set. an honest bet for multi category predictions furthermore. although it needs conditional independence assumption, Naïve Thomas Bayes Classifier has bestowed smart performance in varied application domains.

4.3 Ordinary Least Squares Regression: Least squares may be a methodology for performing simple regression. This can be the task of fitting a line through a group of points. In normal method of least squares strategy one will draw a line, then for every of the info points, live the vertical distance between the purpose and therefore the line, and add these up; the fitted line would be the one wherever this add of distances is as tiny as attainable. Linear refers the sort of model of victimization to suit the info, whereas method of least squares refers to the sort of error metric minimizing over.

Advantages of Linear Regression Machine Learning Algorithm

- It is easy and interpretable machine learning algorithms.
- It requires minimal tuning.
- It is much faster

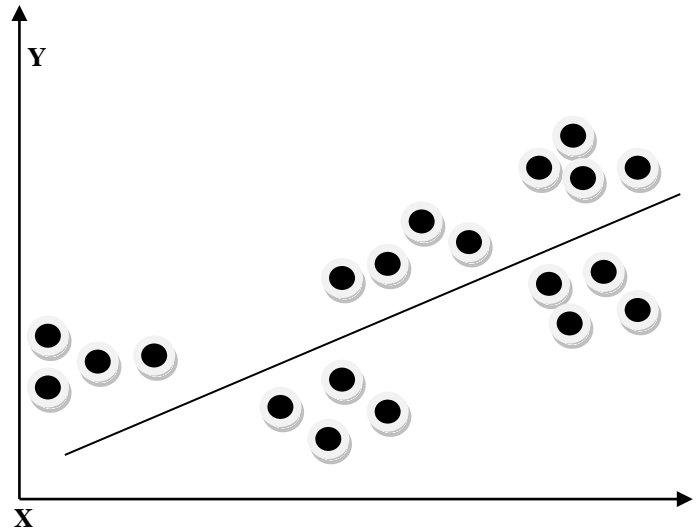


Fig 4.2: Linear Regression

4.4 Logistic Regression: Logistic regression could be a powerful applied mathematical approach of modeling a binomial outcome with one or a lot of informative variables. It measures the connection between the specific variable quantity and one or a lot of freelance variables by estimating chances employing a provision perform, that is that the additive provision distribution. In general, regressions may be employed in real-world applications such as:

- Credit grading
- Measuring the success rates of selling campaigns
- Predicting the revenues of an exact product
- Is there reaching to be Associate in Nursing earthquake on a specific day

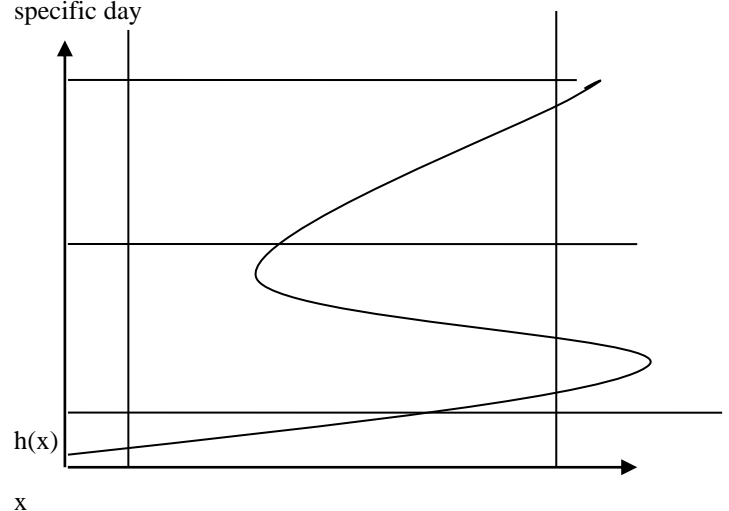


Fig 4.4: Logistic Regression

Advantages of Logistic Regression

Easier to look at and fewer difficult durable rule as a result of the freelance variables needn't have equal variance or distribution. These algorithms do not assume a linear relationship between the dependent and freelance variables and thus may handle non-linear effects. Controls contradictory and tests interaction. Logistic regression is known as once the transformation operate it uses, that is named the logistical operate

$$h(x) = 1 / (1 + e^{-x})$$

This forms a sigmoid curve. In logistic regression, the output takes the shape of chances of the default category (unlike regression, wherever the output is directly produced). Because it may be a chance, the output lies within the variation of 0-1. So, as an example, if we are making an attempt to predict whether or not patients are sick, we have a tendency to already recognize that sick patients are denoted as one, thus if our algorithmic program assigns the score of zero.98 to a patient, it thinks that patient is sort of seemingly to upchuck.

This output (y-value) is generated by log remodelling the x-value, victimisation the logistical operate $h(x) = 1 / (1 + e^{-x})$. A threshold is then applied to force this chance into a binary classification.

Logistic-Function-machine-learning

Logistic Regression to see if a neoplasm is malignant or benign. Classified as malignant if the chance $h(x) \geq 0.5$. Source. To determine whether or not a neoplasm is malignant or not, the default variable is $y = one$ (tumor = malignant). The x variable can be a measuring of the neoplasm, like the scale of the neoplasm. As shown within the figure, the logistical operate transforms the x -value of the assorted instances of the info set, into the vary of zero to one. If the chance crosses the edge of 0.5 (shown by the horizontal line), the neoplasm is classed as malignant.

The logistical equation $P(x) = e^{(b_0 + b_1x)} / (1 + e^{(b_0 + b_1x)})$ is reworked into $\ln(p(x) / 1-p(x)) = b_0 + b_1x$.

The goal of logistical regression is to use the coaching knowledge to seek out the values of coefficients b_0 and b_1 such it'll minimize the error between the anticipated outcome and therefore the actual outcome. These coefficients are calculable victimisation the technique of most probability Estimation.

4.5. Support Vector Machines:

Support Vector Machine is a binary classification algorithm. In this algorithm a set of points of 2 types is given in N dimensional place. This algorithm generates a (N-1) dimensional hyper plane to separate those points into 2 groups. Some of the biggest problems that can be solved with the help of Support Vector Machine are display advertising, human splice site recognition, image-based gender detection, large-scale image classification etc. The similarity between instances is calculated using measures such as Euclidean distance and Hamming distance.

Advantages of Using SVM

- 1.Support Vector Machine offers best classification performance (accuracy) on the training data.
2. SVM renders more efficiency for correct classification of the future data.
- 3.The best thing about SVM is that it does not make any strong assumptions on data.

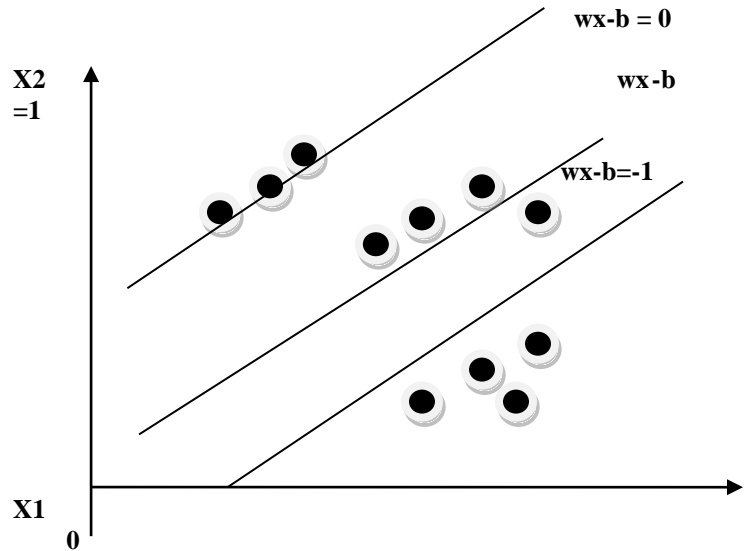


Fig 4.5: Support Vector Machine

Unsupervised learning

Unsupervised learning is enforced once we have less info concerning objects; during this case the coaching set of knowledge is unlabeled. Our goal is to look at some similarities between the teams of the objects and embrace them in acceptable clusters. Some objects will disagree vastly from all clusters; these objects area unit declared because the anomalies.

4.6 Apriori

The Apriori rule is employed in an exceedingly transactional information to mine frequent item sets then generate association rules. it's popularly utilized in market basket analysis, wherever one checks for mixtures of merchandise that regularly co-occur within the information. In general, we tend to write the association rule for 'if an individual purchases item X, then he purchases item Y' as : $X \rightarrow Y$.

Example: if an individual purchases milk and sugar, then he's seemingly to buy occasional powder. this might be written within the kind of AN association rule as: \rightarrow occasional powder. Association rules area unit generated when crossing the edge for support and confidence. The Support live helps prune the amount of candidate itemsets to be thought of throughout frequent itemset generation. This support live is target-hunting by the Apriori principle. The Apriori principle states that if AN itemset is frequent, then all of its subsets should even be frequent.

$$\text{Support} = \frac{\text{freq}(X, Y)}{N}$$

$$\text{Rule: } X \Rightarrow Y \rightarrow \text{Confidence} = \frac{\text{freq}(X, Y)}{\text{freq}(X)}$$

$$\text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)}$$

Fig 4.7: Formulae for support, confidence and lift for the association rule X->Y.

The Support vector helps prune the quantity of candidate item sets to be thought-about throughout frequent item set generation. This support live is guided by the Apriori principle. The Apriori principle states that if AN itemset is frequent, then all of its subsets should even be frequent.

Advantages of Apriori formula

- It is straightforward to implement and may be parallelized simply.
- Apriori implementation makes use of enormous item set properties.

4.7. K-means

K-means is AN reiterative formula that teams similar knowledge into clusters. It calculates the centre of mass of k clusters and assigns a knowledge purpose to it cluster having least distance between its centroid and therefore the datum.

Step 1: k-means initialization:

- opt for a worth of k. Here, allow us to take k=3.
- Arbitrarily assign every datum to any of the three clusters.
- Work out cluster centre of mass for every of the clusters. The red, blue and inexperienced stars denote the centroids for every of the three clusters.

Step 2: Associating every observation to a cluster:

Reassign every purpose to the nighest cluster centre of mass. Here, the higher five points got assigned to the cluster with the blue color centre of mass. Follow a similar procedure to assign points to the clusters containing the red and inexperienced color centre of mass.

Step 3: Recalculating the centroids:

Calculate the centroids for the new clusters. The previous centroids area unit shown by grey stars whereas the new centroids area unit the red, inexperienced and blue stars.

Step 4: Ingeminate, then exit if unchanged.

Repeat steps 2-3 till there's no change of points from one cluster to a different. Once there's no change for two consecutive steps, exit the k-means formula.

Advantages of K-Means Machine Learning formula

just in case of orbicular clusters, K-Means produces tighter clusters than hierarchal bunch. Given a smaller worth of K, K-Means bunch computes quicker than hierarchal bunch for giant variety of variables.

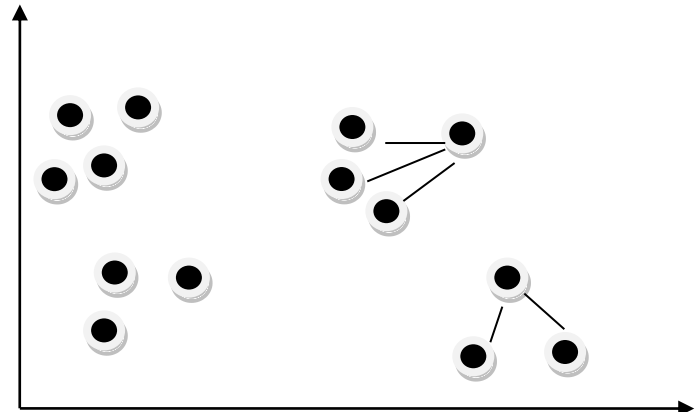


Fig 4.7: Steps of the K-means algorithm.

Here’s how it works: we have a tendency to begin by selecting a worth of k. Here, allow us to say k = three. Then, we have a tendency to willy-nilly assign every information to any of the three clusters. cipher cluster centre of mass for every of the clusters. The red, blue and inexperienced stars denote the centroids for every of the three clusters. Next, designate every purpose to the nearest cluster centre of mass. within the figure higher than, the higher five points got assigned to the cluster with the blue centre of mass. Follow constant procedure to assign points to the clusters containing the red and inexperienced centroids. Then, calculate centroids for the new clusters. The recent centroids area unit grey stars; the new centroids area unit the red, green, and blue stars. Finally, repeat steps 2-3 till there's no change of points from one cluster to a different. Once there's no change for two consecutive steps, exit the K-means rule.

4.9. PCA(Principal element Analysis)

Principal element Analysis (PCA) is employed to create information straightforward to explore and visualize by reducing the quantity of variables. this can be done by capturing the most variance within the information into a replacement co-ordinate system with axes referred to as ‘principal components’. every element may be a linear combination of the initial variables and is orthogonal to 1 another. Orthogonality between parts indicates that the correlation between these parts is zero. the primary principal element captures the direction of the most variability within the information. The second principal element captures the remaining variance within the information however has variables unrelated with the primary element. Similarly, all serial principal parts (PC3, PC4 so on) capture the remaining variance whereas being unrelated with the previous element.

4.9. Semi-supervised learning

Semi-supervised learning rule includes each tagged and unlabelled information. This technique permits to enhance accuracy considerably, as a result of one will use unlabelled information within the toy with alittle quantity of tagged information conjointly. issues wherever you have got an oversized quantity of input file (X) and just some of the info is labelled (Y) area unit referred to as semi-supervised learning issues. These issues sit in between each supervised and unattended learning. an honest example may be a image archive wherever just some of the pictures area unit labelled,

(e.g. dog, cat, person) and also the majority area unit unlabelled. several universe machine learning issues fall under this space. is} as a result of it can be costly or long to label information because it could need access to domain consultants. Whereas unlabelled information is reasonable and straightforward to gather and store. You can use unattended learning techniques to get and learn the structure within the input variables. you'll be able to conjointly use supervised learning techniques to create best guess predictions for the unlabelled information, feed that information into the supervised learning rule as coaching information and use the model to create predictions on new unseen information.

4.10 Reinforcement learning

Reinforcement learning is a very important space of machine learning that involved with however software system agents have to be compelled to take actions in some specific surroundings to maximise some notion of accumulative reward. A Table is unbroken within the memory to recollect the past actions. Reinforcement learning may be a kind of machine learning rule that permits the agent to determine the simplest next action supported its current state, by learning behaviours that may maximize the reward. Reinforcement algorithms typically learn best actions through trial and error. they're usually utilized in artificial intelligence – wherever a automaton will learn to avoid collisions by receiving feedback once bumping into obstacles, and in video games – wherever trial and error reveals specific movements that may increase a player's rewards. The agent will then use these rewards to grasp the best state of game play and opt for future action. Reinforcement algorithms typically learn best actions through trial and error. Imagine, for instance, a computer game during which the player must move to sure places at sure times to earn points. A reinforcement rule enjoying that game would begin by moving willy-nilly however, over time through trial and error, it'd learn wherever and once it required to maneuver the in-game character to maximise its purpose total.

V. MACHINE LEARNING IN HOTEL REVIEWS

Purpose of our research work is to derive the right conclusion about the already rated reviews on various well defined parameters, In this case we may easily analyse the reviews given and can categories into groups and can predict the good or bad reviews using sentimental classification methods. This is the case of supervised learning.

In case of pre-specified parameters are not provided but some pre-defined clusters may be identified and hotel reviews may be fit in some group. The reviews that did not fit any group will be termed as the spam. This is the case of unsupervised learning.

In hotel reviews some clusters and some predictions both may be defined as per the user requirements, and analysis of reviews may be made. This is the case of semi-supervised learning.

In case we are considering and analyzing the environment and outcomes of hotel reviews may generate specific reactions on specific actions then re-enforcement learning must be used for hotel reviews.

It is concluded that appropriate machine learning techniques can be used for analysis of online reviews of the hotels.

REFERENCES

- [1] Visani C. and Jadeja N., (2017), "A Study on Different Machine Learning Techniques for Spam Review Detection", IEEE Transaction 978-1-5386, pp. 1887-1892.
- [2] Xue H., Li F., Seo H. and Pluretti R., (2015), "Trust-Aware Review Spam Detection", IEEE Computer Society Trustcom/BigDataSE/ISPA, pp. 726-733.
- [3] Jiang M. and Cui P., (2016), "Suspicious Behaviour Detection: Current Trends and Future Directions", IEEE Intelligent systems/1541-1672, Computer Society, pp. 31-39.
- [4] Rout J.K., Dalmia A. and Choo K. K. R., (2017), "Revisiting Semi-supervised Learning for Online Deceptive Review Detection", IEEE Access, 2169-3536, pp. 11-19.
- [5] Deng X. and Chen R., (2014), "Sentiment Analysis Based Online Restaurants Fake Reviews Hype Detection", proceedings of APWeb Workshops, Springer International Publishing Switzerland, pp. 1–10.
- [6] Ruchansky N., Seo S. And Liu Y., (2017), "A Hybrid Deep Model for Fake News Detection", Computer Society of India ACM. 978-1-4503-4918-5/17/11, pp. 17- 27.
- [7] Mukherjee A., Venkataraman V., Liu B., Glance N., (2013), "Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews", Technical Report, Department of Computer Science (UIC-CS-2013-03).University of Illinois at Chicago, pp. 268- 279.
- [8] Kokate S., Tidke B., (2015), "Fake Review and Brand Spam Detection using J48 Classifier", International Journal of Computer Science and Information Technologies ISSN: 0975-9646, Vol. 6 (4), pp. 3523-3526.
- [9] Chaitanya Kale, Dadasaheb Jadhav., Tushar Pawar., (2016), "Fake Spam review detection using natural language processing techniques", International journal of innovations engineering research and technology ISSN: 2394-3696, Vol. 3, Issue 1, Jan.-2016, pp. 31-37.
- [10] Adike R.G., Reddy V., (2016), "Detection of Fake Review and Brand Spam Using Data Mining Technique", International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 07; July - 2016 [ISSN: 2455-1457], pp. 251-256.
- [11] Bonde Y.P., Kharabi K.L., Sabale A.N., (2017), "Detection and Elimination of Fake Review from Real- Time Data using Cloud Computing", International Journal of Advance Engineering and Research Development ISSN: 2348-6406 Volume 4, Issue 5, May-2017, pp. 187-194.
- [12] Crawford M., Khoshgoftaar T.M., Prusa J.D., Richter A.N. and Najadain H.A., (2015), "Survey of review spam detection using machine learning techniques", Springer Journal of Big Data Machine Learning Methods Crawford et al, pp. 17- 39.
- [13] Elmurngi E. and Gherbi A.,(2017), "An Empirical Study on Detecting Fake Reviews", proceeding of IEEE The Seventh International Conference on Innovative Computing Technology, pp. 107-114.
- [14] Fontanarava J., Pasi G. and Viviani M., (2017), "Feature Analysis for Fake Review Detection through Supervised Classification", proceedings of IEEE International Conference on Data Science and Advanced Analytics, pp. 658-666.
- [15] Wahyuni E. D. And Djunaidy A., (2017), "Fake review detection from a product review using modified method of iterative computation framework", proceedings of MATEC Web of Conferences, pp. 121-127.
- [16] Lin Y., Zhu T., Wu H., Zhangl J., Wang X., Zhou A., (2014), "Towards Online Anti-Opinion Spam: Spotting Fake Reviews from the Review Sequence", proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 261-264.
- [17] Li Y., Feng X., Zhang S., Li Y., (2016), "Detecting Fake Reviews Utilizing Semantic and Emotion Model", proceedings of 3rd IEEE-International Conference on Information Science and Control Engineering, pp. 317-320.
- [18] Yin R., Wang H., Liu L. , (2015), "Research of Integrated Algorithm", proceedings of 4th IEEE International Conference on Computer Science and Network Technology, pp. 584- 589.
- [19] Rajamohana S.P., Umamaheswari K., Dharani M., Vedackshya R., (2017), "A Survey on online review spam detection techniques",

proceedings of IEEE International Conference on Innovations in Green Energy and Healthcare Technologies, pp. 8- 13.

- [20] Liu P., Xu Z., Ai J. , Wang F., (2017), “Identifying Indicators of Fake Reviews Based on Spammer’s Behavior Features”, proceedings of IEEE International Conference on Software Quality, Reliability and Security, pp. 396- 403.
- [21] Chauhan S.K., Goel A., Goel P., Chauhan A. and Gurve M.K., (2017), “Research on Product Review Analysis and Spam Review Detection”, proceedings of IEEE 4th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 399- 393.
- [22] Christopher S.L. and Rahulnath H. A., (2016), “Review authenticity verification using supervised learning and reviewer personality traits”, proceedings of IEEE International Conference on Emerging Technological Trends, pp. 16- 23.
- [23] Ohana B and Tierney B, “Sentiment classification of reviews using SentiWordNet” , 9th. IT & T Conference. 2009: pp 1232-1243

AUTHORS PROFILE



Pankaj Chaudhary has completed his B.Tech and M.Tech, he is Ph.D(CSE) reasearch scholer at ICFAI University, Dehradun . He has published 16 National and International research papers in journals of repute. He has also attended several conferences.Currently he is doing his reasearch in analysing the genuinity of the online reviews of hotels.



Dr. Anurag Aeron is Associate professor(CSE) at ICFAI University, Dehradun, He has completed his Ph.D from IIT Roorkee. His research areas are Remote Sensing and GIS, Open Source Systems, Disaster Management, AI, Android Operating System, Machine Learning, IOT, NLP



Dr. Sandeep Vijay, Working as Director at Shivalik College of Engineering, Dehradun, He has completed his Ph.D from IIT Roorkee. He has Proficiency in spearheading overall strategic research and developments projects, right from planning, cost controls,, resource mobilization, structured communications to final reviews, within cost & time parameters.