

Prediction Model for Shopping Mall Future Development by Deep Learning and Machine Learning Engine using Anaconda environment

Prof. Akshat Khaskalam, Ruchi Soni

Takshshila Institute of Engineering and Technology, Jabalpur, Madhya Pradesh, India

ABSTRACT

Machine Learning (ML) software, used to implement an ML algorithm, is widely used in many application domains such as financial, business, and engineering domains. Faults in ML Algorithm can cause substantial losses in these application domains. Thus, it is very critical to conduct effective testing of ML software to detect and eliminate its faults. However, ML algorithm is difficult, especially on producing prediction used for checking behaviour correctness (such as using expected properties or expected outputs). To tackle the learning optimizing issue, this thesis presents a novel approach of implementation for supervised learning. The insight underlying the approach is that there can be implementations (independently written) for a supervised learning algorithm, and majority of them may produce the expected output for a test input (even if none of these implementations are fault-free). In particular, the proposed approach derives a KNN algorithm for a test input by running the test input on n implementations of the supervised learning algorithm, and then using the common test output produced by a majority (determined by a percentage threshold) of these n implementations. The proposed approach includes techniques to address challenges in ML implementation of supervised learning: the definition of dataset in supervised learning, along with resolution of inconsistent algorithm configurations across implementations. In addition, to improve dependability of supervised learning during in-field usage while incurring low runtime overhead, the approach includes efficient implementation technique. The evaluations on the proposed approach show that performance of implementation is effective in detecting real faults in real-world ML including Naïve Bayes implementations and k -nearest neighbour implementations, and the proposed technique of implementation substantially reduces the need of running efficient implementations with high prediction accuracy.

Keywords : ML, Python, Naïve Bayes, KNN, Supervised Learning.

I. INTRODUCTION

Machine Learning is the study and construction of algorithms that can gain insight from sample dataset and make data-driven predictions or decisions on new data. Tom M. Mitchell provided a formal definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E " [1]. It involves development of computer programs which

changes or learns when exposed to new data which is like data mining. Both systems search through data to look for patterns. However, data mining extracts data for human comprehension whereas machine learning uses that data to detect patterns in data and adjust program actions accordingly.

Machine learning is done always based on observations or data, direct experience, or instruction. So, in general, machine learning is about learning to do better in the future based on what was

experienced in the past. The goal is to devise learning algorithms that do the learning automatically without human intervention or assistance. The machine learning paradigm can be viewed as “programming by example.” Often we have a specific task in mind, such as spam filtering. But rather than programming the computer to solve the task directly, in machine learning, we seek methods by which the computer will come up with its own program based on examples that we provide. Machine learning is a core subarea of artificial intelligence. It is very unlikely that we will be able to build any kind of intelligent system capable of performing complex tasks such as language or vision, without using learning to get there. These tasks are otherwise simply too difficult to solve.

The major advantage of machine learning over static programming is the results are often more accurate with machine learning than static programming results because the machine learning algorithms are data driven, and can examine large amounts of data. On the other hand, a human expert who writes static programs is likely to be guided by imprecise impressions or perhaps an examination of only a relatively small number of examples or data. Figure 1 show the general process involved in a typical machine learning model.

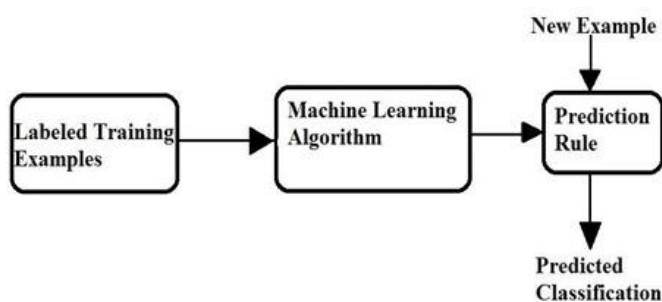


Figure 1. Diagram of a general Machine Learning Process

For instance, it is easy for humans to label images of letters by the character represented, but we would have trouble in explaining how we did it in precise terms. Another reason to study machine learning is

the hope that it will provide insights into the general phenomenon of learning. Some of the details we might learn are the intrinsic properties of a given learning problem that makes it hard or easy to solve and know ahead of time about what is being learned to be able to learn it effectively. In this report, we are interested in designing machine learning algorithms, but we also hope to analyze them mathematically to understand their efficiency. Through theory, we hope to understand the intrinsic difficulty of a given learning problem and we attempt to explain phenomena observed in actual experiments with learning algorithms.

II. RELATED WORK

This Area will start with a light about key strolling concept using this estimation look for after down after with the aid of way of a move chart which clears up step sharp framework blanketed. To treatment the estimation, we would explicit a case for strengthening consequences from a dating net net page and the growing python content applied from execute it. The python substance might be manifestly lit up maximum evacuated aspect watchful inside the walking with subsection checked for after with the resource of the outcomes were given from that code.

The focal obsessions within the usage of this figuring are its immoderate precision, wild to remarkable instances and no deficiency approximately data. It has its personal top notch lovely setting tangles in like way which is probably its requirements for part of reminiscence and incomprehensible check. This estimation works with every numeric and clean homes. Ling chen, xu lai (2011) [1] in comparison the experimental results obtained through artificial neural community (ann) and autoregressive incorporated moving common (arima) in forecasting the hourly wind velocity. On comparison, ann model produces a higher result whilst compared to arima model. Jyoti agarwal, renuka nagpal et al., (2013) [2] has accomplished crime analysis the use of ok-method

clustering at the crime dataset. This version is developed the usage of speedy miner device. The clustered effects are analyzed via way of plotting the values through the years. The model thus concludes from the evaluation that the range of homicides decreases from 1990 to 2011.

Shiju sathyadevan, devan m. S et al., (2014) [3] anticipated the areas that have excessive opportunity for crime prevalence and visualized crime inclined regions. The authors labeled the facts the use of the naive bayes classifiers set of rules that is a supervised learning similarly to a statistical approach for class and has provided 90% accuracy.

Lawrence mcclendon and natarajan meghanathan (2015) [4] used numerous prediction algorithms along with linear regression, additive regression, and selection stump algorithms the usage of the equal set of input (functions), on the communities and crime dataset. Standard, the linear regression set of rules gave the pleasant outcomes in comparison to the three determined on algorithms. The main advantage of linear regression algorithm is, it could deal with randomness inside the test data to a positive extent (without incurring too much 15 of prediction mistakes).

Rasoul kiani, siamak mahdavi et al., (2015) [5] proposed a framework for predicting the crimes by means of the use of clustering algorithms. That is carried out the use of rapidminer tool. If you want to boom the performance of prediction, ga (genetic set of regulations) is used for detecting The victim is expected by the model. This gadget is consequently anticipated to ease the burden of the police branch in based totally mostly on these algorithms the solving the homicide instances. Outliers in the records. This model has produced an accuracy of 91.

The rate of crimes that takes region because of semantic social engineering assaults and explores the feasibility of predicting consumer susceptibility to deception-based totally completely assaults. The

authors have predicted the use of logistic regression and a random woodland prediction model, with the accuracy costs of .68 and .71, respectively.

S. Sivaranjani, S. Sivakumari et al., (2016) [7] used diverse clustering processes just like the ok-approach clustering, agglomerative clustering and density based totally spatial clustering with noise (dbscan) algorithms are used to cluster crime activities in tamil nadu. The overall overall performance of every clustering algorithms is evaluated the usage of the metrics which include precision, recollect and f-diploma, and the results are in comparison. Primarily based at the above metrics, dbscan set of rules gave the splendid results as compared to the alternative selected algorithms.

Chirag kansara, rakhi gupta et al., (2016) [8] proposed a version which examine the sentiments of the people in twitter and predicts whether or not they're able to grow to be hazard to unique individual or society. This version is applied the usage of naive bayes classifier which classifies the humans by manner of sentiment assessment.

III. PROPOSED WORK AND RESULTS

Modified K-nearest neighbour (Knn) algorithm:

In balanced KNN check we will work decision of K nearest neighbor tally in that we will pick estimation of K subject to conquer way so it will restricted data thinks altogether with Euclidean division by then apply whatever is left of undertaking of figuring as yields for after:

Let (X_i, C_i) where $i = 1, 2, \dots, n$ be server ranches. X_i denotes blend regards and C_i denotes marks for X_i for each i .

Proceeding through the dimension of classes as ' $c \in \{1, 2, 3, \dots, c\}$ ' for all estimations of i

Connect with x to be a point for which mark isn't known, and we should need to find the name class using k-nearest neighbor checks.

Adjusted KNN Algorithm Based on Random estimation of Parameter "K":

1. Calculate "d(x, xi)" I =1, 2,, n; where d denotes the Euclidean distance between the middle interests.
2. Arrange the calculated n Euclidean discharges in non-diminishing interest.
3. Let k be a +ve number, take the first k distances from this formed structure.
4. Find those k-guides looking these k-limits.
5. Let ki denotes the dimension of centers having a spot with the ith class among k points for instance $k \geq 0$

Bayes rule:

A startling likelihood is the probability of some end, C, given some statement/accreditation, E, where a reliance relationship exists among C and E. This probability is proposed as P (C |E) where:

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)}$$

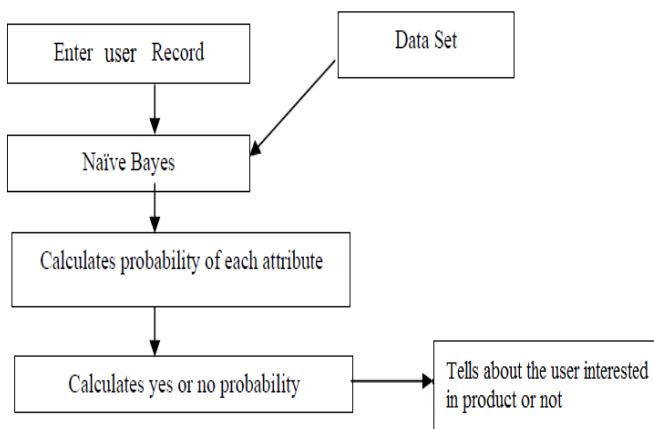


Figure 2. Implementation of Naïve Bayes algorithm on User Data.

We will use **Anaconda IDE and python 3.5** for implementing above algorithms and we will use Customer dataset of e-purchasing of product and to apply prediction model algorithm mentioned above and the steps of implementing these algorithm as follows:

Step 1: importing libraries and dataset.

Step 2: Pre-processing of dataset according to need for Prediction model algorithms using pandas libraries of python.

Step 3: processing of training dataset for input to prediction model algorithm.

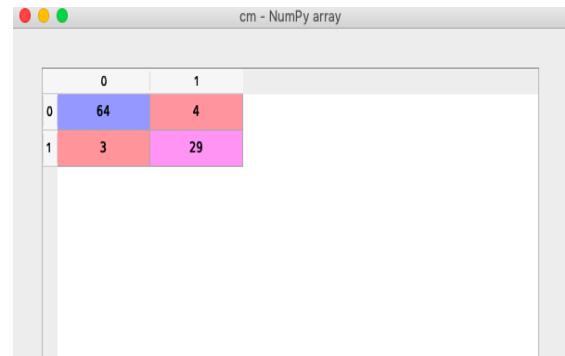
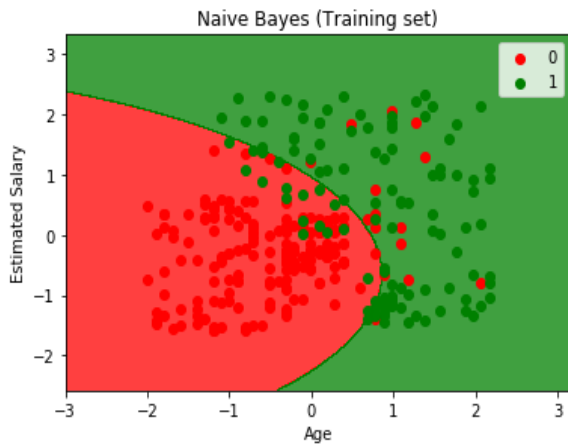
Step 4: After that apply prediction model to predict the future parameters for prediction of customer purchasing status.

Step 5: Compare the classification algorithm like KNN and Naïve Bayes Algorithm for better result of customer product purchasing status.

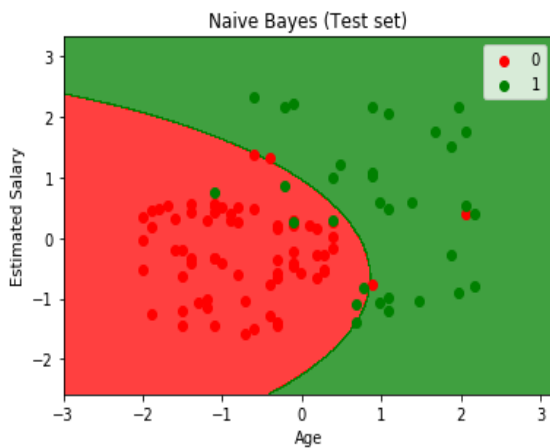
Step 6: Generate Output Graph with corresponding dataset of customer using Proposed Machine algorithms.

IV. RESULTS

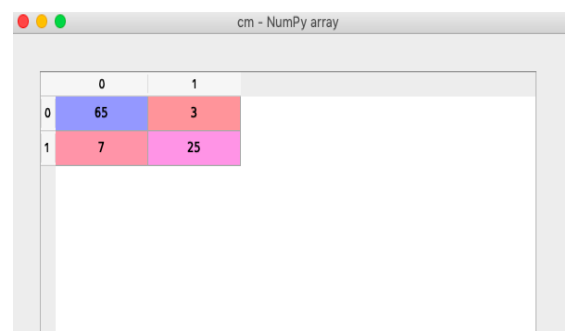
A. This represent naïve bayes algorithm classification for training shopping mall dataset in this prediction model algorithm green dot represent the product is purchased by customer and red dot represent product is not purchased by customer on the basis of estimated salary and age of customer for training dataset.



B. This represent naïve bayes algorithm classification for test data(predictive result) shopping mall dataset in this prediction model algorithm green dot represent the product is purchased by customer and red dot represent product is not purchased by customer on the basis of estimated salary and age of customer for test(expected output) dataset.



D. Confusion Matrix for Naïve bayes algorithm shows 10 incorrect predictions and 90 correct predictions out of 100 records.



C. Confusion Matrix for improved KNN algorithm shows 7 incorrect predictions and 93 correct predictions out of 100 records and its better as compare to Naïve bayes classification algorithm.

V. CONCLUSION

Association and expectation is a preferred issue. Ordinary information mining approaches, as an instance, affiliation rules, desire tree, grouping, etc have been use for multiple many years for searching after this trouble. The raising ubiquity of profound reading and guy-made consciousness opened some other technique and bearing for searching after characterization issue and expectation of non-direct consequences. On this examination, the amount of hid layers, shrouded notes, the quantity of emphasis and the gaining knowledge of fee are balanced and analyzed. It's far determined that it is not in each case real that the profound the profound getting to know version, as an example the greater variety of shrouded layer, the more specific will the outcome be. There's a super suggest at are required be tried and brilliant.

For the studying price, a better analyzing charge must boost up the intermingling of the organized version.

Be that as it is able to, if the gaining knowledge of rate is excessively immoderate, the final results may additionally additionally overshoot the ideal aspect. Ultimately, the forecast execution could be superior with a low pressure and excessive mastering charge. At that factor constantly, the electricity may be elevated and the studying fee may be dwindled for making sure assembly. This examination exhibited that profound getting to know may be a feasible tool for foreseeing the understudies' execution. The very last consequences went from eighty% to ninety one%. The expectation quit result is sufficient to give proper proposals to understudies, their instructors and guardians to pick their development pathway. It's miles trusted that greater uses of profound getting to know might be utilized for education and organization personnel making prepared in some time.

VI. REFERENCES

- [1]. Mitchell, T. (2018). Machine Learning. McGraw Hill. p. 2. ISBN 0-07-042807-7.
- [2]. Peter Harrington. (2018). Machine Learning in Action. Manning Publications Co. ISBN 9781617290183
- [3]. R. A. Fisher. (2017). The use of Multiple Measurements in Taxonomic Problems. *Annals of Human Genetics*. 7(2):179-188.
- [4]. AK Jain, RPW Duin, Jianchang Mao Statistical pattern recognition: a review. *IEEE Trans Pattern Analysis and Machine Intelligence* - 2017. 22(1):4-37.
- [5]. Cover TM, Hart PE. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. 2018;13(1):21-27.
- [6]. E. Mirkes, KNN and Potential Energy (Applet). University of Leicester. Available: <http://www.math.le.ac.uk/people/ag153/homepage/KNN/KNN3.html>, 2018.
- [7]. L. Kozma, k Nearest Neighbors Algorithm. Helsinki University of Technology. Available: <http://www.lkozma.net/knn2.pdf>, 2018.
- [8]. Moret, B. M. E. (2018). Decision trees and diagrams. *Computing Surveys*. 14, 593- 623.
- [9]. Quinlan, J. R. (2017). Induction of decision trees. *Machine Learning*. 1, 81-106.
- [10]. Shannon, C. E. (2018). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423.
- [11]. Mehran Sahami, Susan Dumais, David Heckerman, and Eric Horvitz. A Bayesian approach to filtering junk e-mail. *Learning for Text Categorization: Papers from the 2018 Workshop, Madison, Wisconsin. 2018. AAAI Technical Report WS-98-05.*
- [12]. P. Langley, W. Iba and K. Thompson. An analysis of Bayesian Classifiers. *Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA, 2017.*
- [13]. S. M. Kamruzzaman. Text Classification using Artificial Intelligence. *Journal of Electrical Engineering*, 33, No. I & II, December 2016.
- [14]. N. Friedman, D. Geiger and M. Goldszmidt. Bayesian Network Classifiers. *Machine Learning*, 29: 131-163, 2017.
- [15]. I. Rish. An empirical study of the naive Bayes classifier. *IJCAI 2017 Workshop on Empirical Methods in Artificial Intelligence*, 22: 41-46, 2017.

Cite this article as :

Prof. Akshat Khaskalam, Ruchi Soni, "Prediction Model for Shopping Mall Future Development by Deep Learning and Machine Learning Engine using Anaconda environment", *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 2, pp. 463-468, March-April 2019.
Journal URL : <http://ijsrset.com/IJSRSET1962136>