An Empirical Comparison of Naïve Bayes and Multinomial Naïve Bayes Classification

Chandra Sekar K[1], Anjan Babu G[2]

Dept of Computer Science, SV University, Tirupati

Abstract— *Ongoing work in managed learning has shown that a shockingly basic Bayesian classifier with solid suppositions of autonomy among highlights, called guileless Bayes. This reality brings up the issue of whether a classifier with less prohibitive suspicions can perform shockingly better. In this paper we assess approaches for prompting classifiers from information, in view of the hypothesis of learning Naïve Bayes and Multinomial Naïve Bayes models. These models are calculated portrayals of likelihood dispersions that sum up the gullible Bayesian classifier and expressly address articulations about autonomy. We tentatively tried these methodologies on Mushroom dataset from the University of California at Irvine storehouse, to identify mushrooms are consumable or noxious.*

I. INTRODUCTION

The objective of assortment learning is to energize a model that isolates the information into the various classes, totally plan on mentioning new models later on. Social occasion learning philosophies rather produce various models. Given another model, the organization passes it to the entirety of its different base models, gets their suspicions, and from that point obliges them in some fitting way (e.g., averaging or projecting a surveying structure). Most of outfit learning methodologies are standard, material across wide classes of model sorts and learning assignments. Organization learning is a reasonable technique that has progressively been embraced to join different learning assessments to furthermore encourage when in doubt figure exactness [3]. Possibly the most novel spaces of examination in supervised AI have been to scrutinize frameworks for making uncommon outfits of understudies. The key openness is that outfits are every so often essentially more definite than the individual understudies [5]. When orchestrating an organization learning technique, similarly as picking the method by which to achieve arrangement in the base models and picking the joining methodology, one necessity to pick the kind of base model and base model learning assessment to utilize. The joining method might limit such base models that can be utilized.

With the fast improvement of data advancement and affiliation headway, various exchanges produce a huge load of information dependably. The real information can't pass on direct advantages so need to sensibly mine covered data from colossal extent of information. Information tunneling directs looking for charming models or information from tremendous information. It changes a massive mix of information into information. Information mining is a focal improvement during the time spent information revelation. The information mining has become an entrancing gadget with respect to breaking down information according to substitute viewpoint and changing over it into significant and gigantic data [8]. Information mining has been all things considered applied in the space of clinical finding, Intrusion recognizing verification framework, Education, Banking, Fraud revelation. Social affair is a controlled learning. Guess and plan in information mining are two sorts of information assessment task that is utilized to detach models depicting information classes or to expect future information plans. Depiction measure has two stages; the first is the learning association where the preparation illuminating documents are examined by get-together assessment. The learned model or classifier is introduced as plan rules or models. The resulting stage is the utilization of model for social occasion, and test edifying arrangements are utilized to study the accuracy of depiction rules.

II. CLASSIFICATION INTERACTION

Over the previous decade there has been an increment in the work done on applying AI calculations to the clinical space. Characterization is a champion among the most examined issues in AI and data mining [5]. Expecting the consequence of a disease is a champion among the most interesting and inciting tasks in which to make data mining applications.

Portrayal is the way toward learning the target limit that guides between a great deal of features and predefined class marks. The data for the gathering is a great deal of events. Every event is a record of data as (X, Y) where X is the features set and Y is the goal variable.

Grouping of this enormous measure of information is tedious and uses unreasonable computational exertion, which may not be suitable for some applications. The order of clinical information has become an inexorably difficult issue, because of late advances in clinical mining innovation. Arrangement focuses on to characterizing a theoretical model of a bunch of classes, called classifier, which is worked from a bunch of marked information, the preparation set. The classifier is then used to fittingly characterize new information for which the class mark is obscure [2][7]. Building precise and productive classifiers for medical data sets is one of the fundamental errands of information mining and AI research. Building viable arrangement frameworks is one of the focal assignments of information mining.

The arrangement data involves events whose class names are known. The course of action model can be developed ward on the arrangement data. The model by then can be surveyed and attempted by using the testing data which contains records with dark class marks.

Grouping is a two-phase measure:

1. **Model turn of events:** portraying a great deal of predestined classes. Each tuple is acknowledged to have a spot with a predefined class, as constrained by the class mark attribute. The plan of tuples used for model turn of events: getting ready set. The model is addressed as gathering rules, decision trees, or logical formulae.
2. **Model use:** for requesting future or dark articles. It measures accuracy of the model; the known name of test is differentiated and the portrayed result from the model. Precision rate is the degree of test set models that are viably requested by the model. Test set is liberated from planning set, for the most part over-fitting will occur.

III. METHODOLOGY

Perhaps the most interesting spaces of examination in oversaw AI have been to scrutinize procedures for building phenomenal social events of understudies.

3.1 Naive Bayes

The Naive Bayes is an energetic procedure for arrangement of quantifiable farsighted models. NB relies upon the Bayesian theory [1][4]. This estimation uses class prohibitive self-rule and has ability to adjust quickly. This portrayal technique assessments the association between every property and the class for every guide to decide a prohibitive probability for the associations between the trademark characteristics and the class [6]. In the midst of setting up, the probability of each class is enrolled by checking how frequently it occurs in the planning dataset. This is known as the "prior probability" P(C=c). Despite the previous probability, the computation also enlists the probability for the event x given c with the assumption that the characteristics are self-governing. This probability transforms into the consequence of the probabilities of each single quality. The probabilities would then have the option to be evaluated from the frequencies of the events in the planning set.

3.2 Bayesian Theorem

Given training data X, posterior probability of a hypothesis H, P(H|X), follows the Bayes theorem $P(H|X) = \frac{P(X|H)P(H)}{P(X)}$

Let X be data tuple and H be some hypothesis such that the data tuple X belongs to a specified class C. For classification problems, we want to determine P (H|X), the probability that the hypothesis H holds the given evidence or observed data tuple X.

P (H|X) is the posterior probability of H conditioned on X

P (H) is the prior probability of H

P (X|H) is the posterior probability of X conditioned on H

P(X) is prior probability of X

3.3 Multinomial Naïve Bayes

Multinomial Naive Bayes is favored to use on data that is multinomial distributed. It is widely used in text classification in NLP [6]. Each event in text classification constitutes the presence of a word in a document.

$$\begin{aligned}\log p(C_k \mid \mathbf{x}) &\propto \log\left(p(C_k)\prod_{i=1}^{n} {p_{ki}}^{x_i}\right)\\ &= \log p(C_k) + \sum_{i=1}^{n} x_i \cdot \log p_{ki}\\ &= b + \mathbf{w}_k^\top \mathbf{x}\end{aligned}$$

IV. EXPERIMENTAL RESULTS

This part gives results and related discussion on data driven investigation of mushroom dataset was accumulated from UCI store [9]. This investigation work was executed using Weka. WEKA is made by investigators at the University of Waikato in New Zealand. The item is written in the Java language and contains a GUI for speaking with data reports. WEKA furthermore gives the graphical UI of the customer and gives various workplaces. WEKA is a state-of-the-art office for making AI (ML) strategies and their application to genuine data mining issues. The Mushroom dataset subtleties are displayed in the table-1.

TABLE 1
MUSHROOM DATASET INFORMATION

Dataset	No. of Attributes	No. of Instances	No. of Classes
Mushroom	23	8124	Eatable-4208, Poison-3916

We utilize 70% of records as the preparation information and the other 30% as the testing information. The Mushroom dataset attribute distribution and statistical summary are shown in the figure-1 and figure-2.

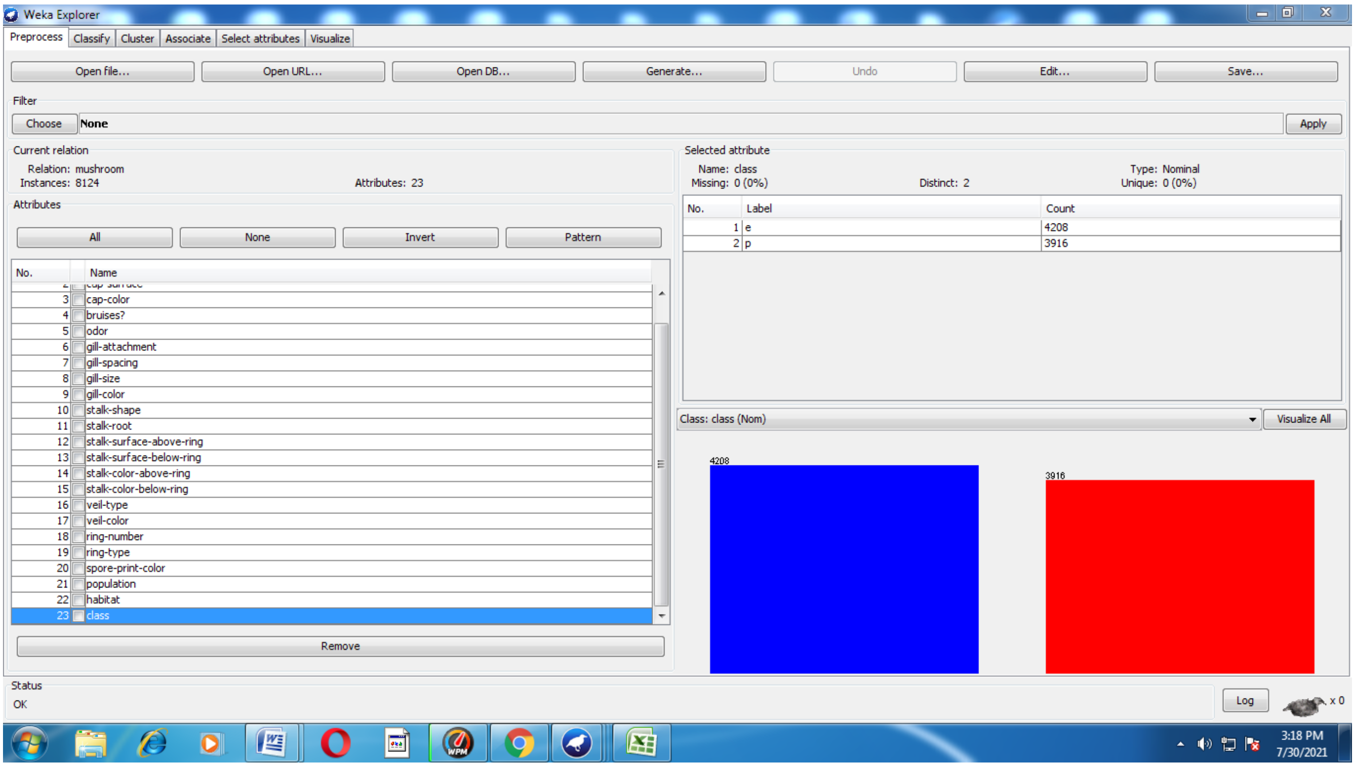

FIGURE 1: Mushroom dataset attribute distribution

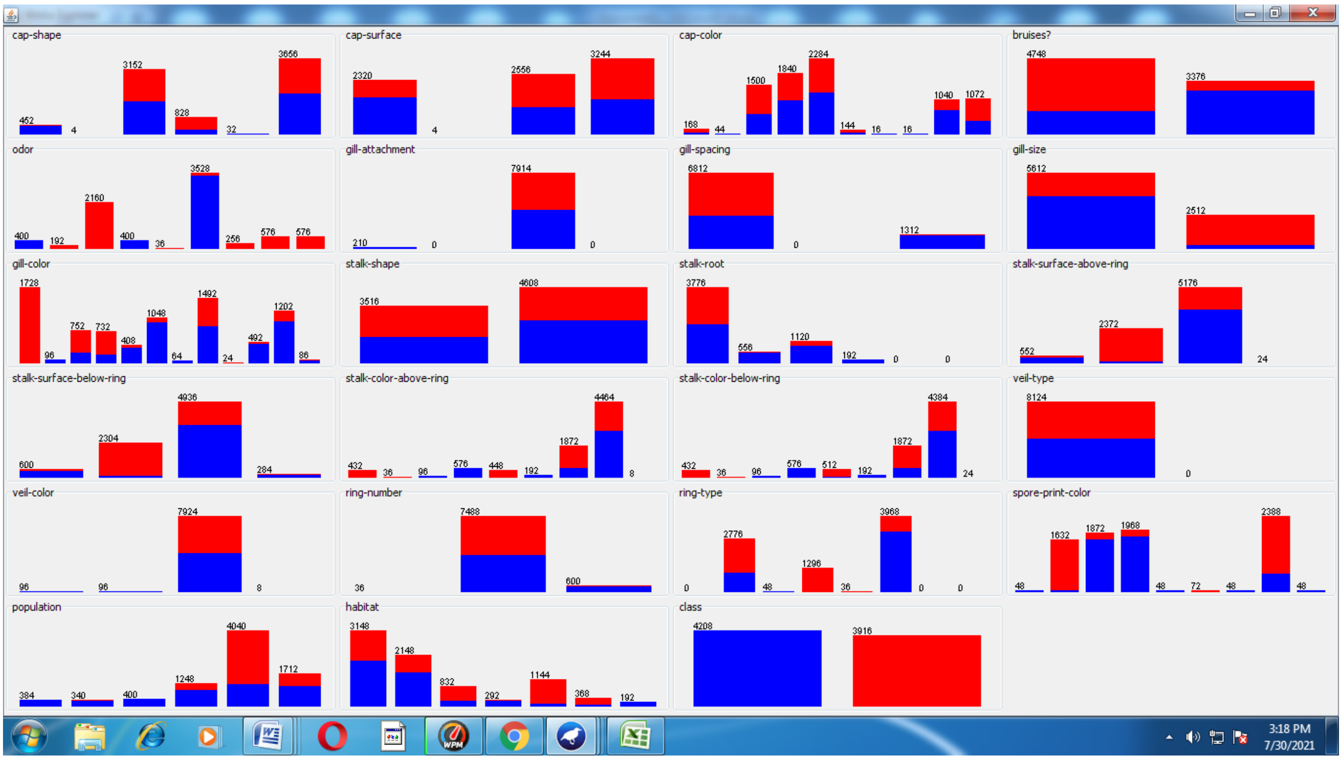

FIGURE-2: Mushroom dataset attribute distribution

4.1 Results

The results of naïve bayes and multinomial naïve bayes classifiers are compared the on basis of correctly classified instances is shown in the table-2 and same shown in the figure-3 with their corresponding values.

TABLE-2
PERFORMANCE OF CLASSIFIERS

Algorithm	Accuracy	precision	Recall
Multinomial Naive Bayes	95.8	96	95.8
Naïve Bayes	96.2	96.2	96.2

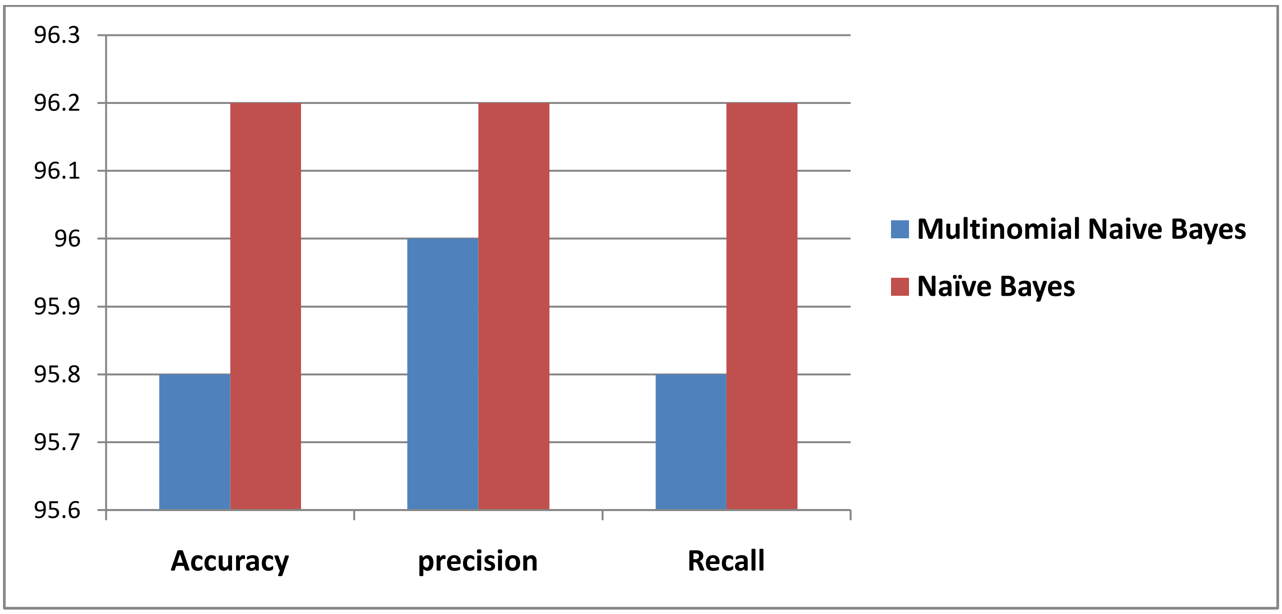

FIGURE 3: Classifier performance

From the figure-3, we notice the exhibition of classification for Multinomial Naive Bayes 95.8% of accuracy and the Naive Bayes has achieved the accuracy of 96.2%. So, the Naive Bayes classification has got highest accuracy when compared to Multinomial Naive Bayes. So, the both algorithms have got highest accuracy, but only 0.4% difference of Naive Bayes when compared to Multinomial Naive Bayes. The screen shots of experimental results are shown in the figure-4 and figure-5.

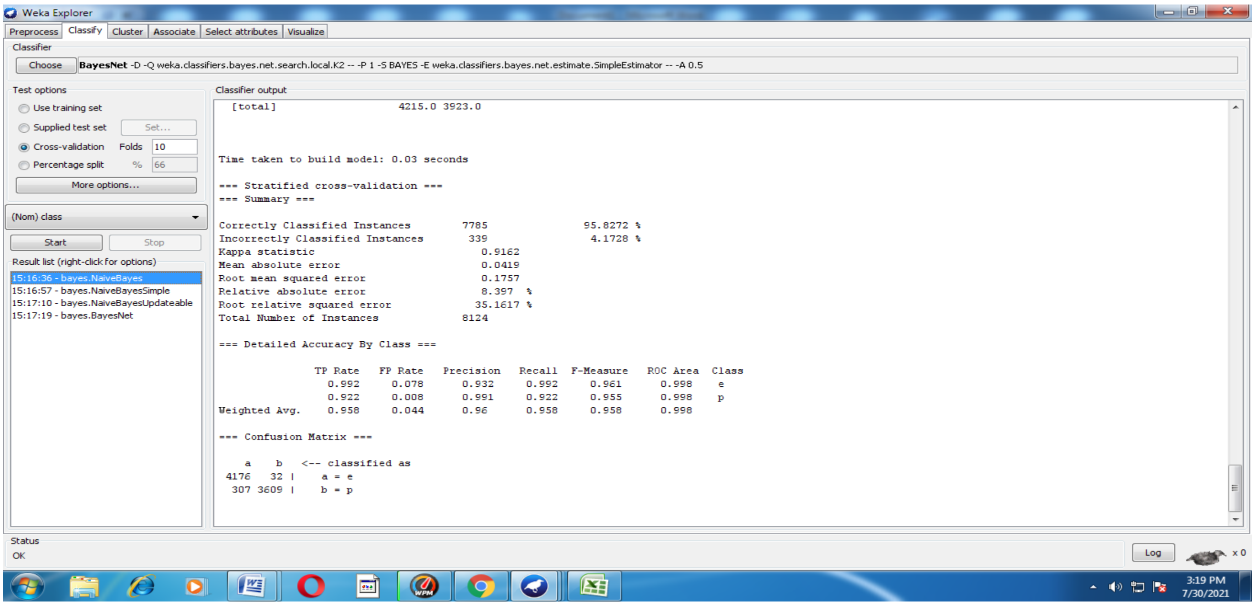

FIGURE 4: screen shot of experimental results

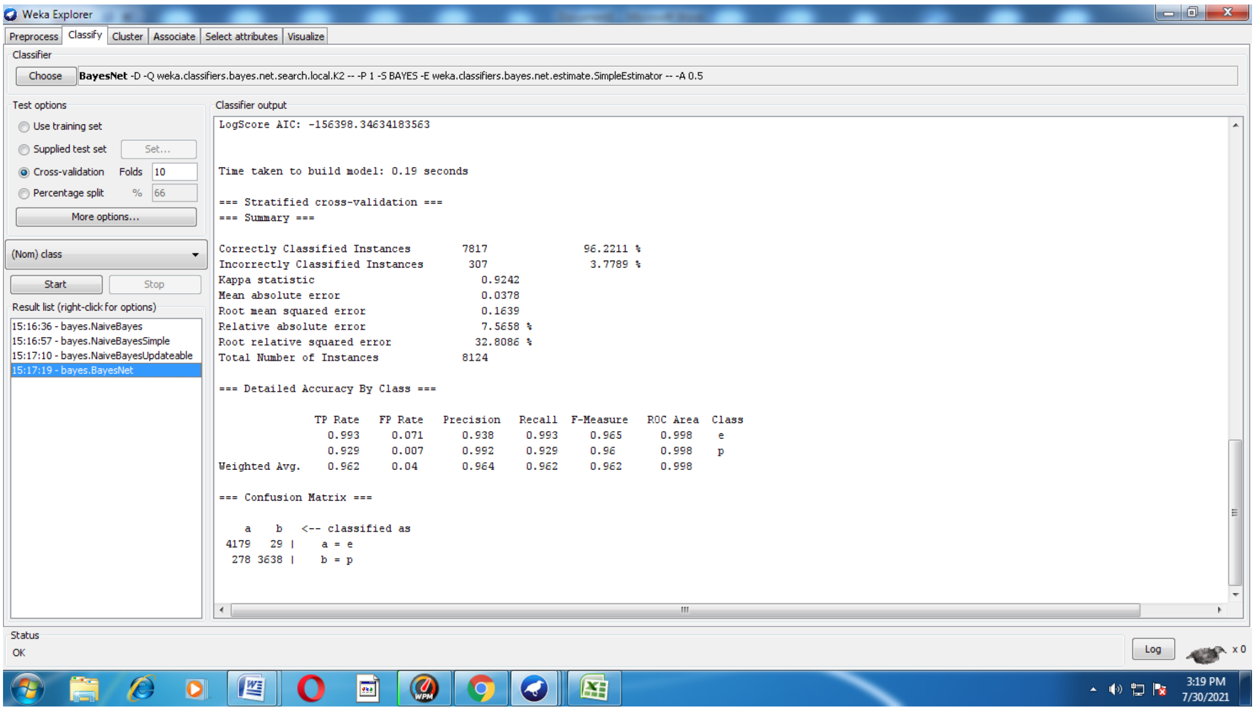

FIGURE-5: screen shot of experimental results

V. CONCLUSION

The Naïve Bayes way to deal with AI is known to have both hypothetical and useful benefits. This paper presents mushroom class location whether mushrooms are palatable or toxic utilizing Naïve Bayes and Multinomial Naïve Bayes models. Our test results show that Naïve Bayes has accomplished most noteworthy exactness on mushroom dataset when contrasted with Multinomial Naïve Bayes. We essentially endeavored to review the turn out achieved for accuracy improvement and execution improvement of Naïve Bayes. This examination which is presented will fill in generally speaking for pursuing future investigation related to self-assertive boondocks classifier.

REFERENCES

[1] Arthur Chol, NazgolTavabi, Adnan Darwiche, ―Structured Features in Naive Bayes Classification‖, Association for the Advancement of Artificial Intelligence, 2016.

[2] G. Ravi Kumar, Venkata Sheshanna Kongara & Dr. G. A. Ramachandra, "An Efficient Ensemble Based Classification Techniques for Medical Diagnosis", International Journal of Latest Technology in Engineering, Management and Applied Sciences, Volume II, Issue VIII, Pages: 5-9, ISSN-2278-2540, August-2013

[3] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques.2nd ed. San Francisco: Morgan Kaufmann, 2005.

[4] I. Rish, J. Hellerstein, and T. Jayram. An analysis of data characteristics that affect naive Bayes performance. Technical Report RC21993, IBM T.J. Watson Research Center, 2001.

[5] J. Han and M. Kamber," Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2nd ed. San Mateo, CA; Morgan Kaufmann, 2006.

[6] N. Friedman, D. Geiger, and Goldszmidt M. Bayesian network classifiers. Machine Learning, 29:131–163, 1997.

[7] N. Michael, "Artificial Intelligence - A Guide to Intelligent Systems",2nd edition, Addison Wesley, 2005.

[8] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Reading, MA: Addison-Wesley, 2005.

[9] UCI machine learning repository. http://archive.ics.uci. edu/ml/.