IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 25-30

Criterion Analysis for Anticipating College Admission

Pratik Patil[1], Nikhil Salunkhe[1], Sujit Mandal[1], Ditixa Vyas[1,] Neha Kunte[1]
[1]*(Computer Engineering, Atharva College of Engineering, Mumbai, India)*

Abstract: *The system is a predictive model for engineering student's college prediction. College Admission Predictor System is a Web based application system within which students needs to enter their HSC and CET marks. Now days, problem becomes more difficult and students fail to understand which college they are likely to get admission even after going through cut-off data of previous years. The system makes use of more number of attributes for more accuracy. Accurate prediction is very important for candidate to fill the application form. Accurate prediction of the performance of college during the student admission process is also important. For this programs applying Naïve bayes and Support vector machine (SVM). This model focuses on the student selection for the university and applies machine learning algorithms to admission dataset.*

I. Introduction

Today, there are many students who are applying for engineering after their H.S.C. In this case it is necessary for the students to know what are their chances of getting an admit from such universities/institutes. And in the same case it is also necessary from the university's perspective to know from the total number of applicants who could get admit based on certain criteria. Currently, universities manually check and count the total number of applicants who could get admit into university. This method is prone to human error and thus accounts for some inaccuracies and is also slow and not very consistent for universities to get an actual result. The Conventional prediction methods which are used for admission prediction waste so much time and money of the student. Because of this new methods or projects has to be proposed. This model focuses on the student selection for the university and applies machine learning algorithms to admission data set. The system is Prediction model for college prediction purpose which uses a data mining technique for the prediction. The dataset has been collected over the particular institute from the period of 2011 to 2016. The dataset is being pre-processed to make the dataset valid.

II. Material And Methods

1.Machinelearning

Table 1: Comparison of Machine-Learning Tools

Parameters	Python	R	Spark	Mat lab	Tensor flow
License	Open source	Open source	Open source	proprietary	Open source
Distributed	NO	No	Yes	NO	No
Visualization	Yes	Yes	No	Yes	No
Neural Nets	Yes	Yes	Multilayer	Yes	Yes
Variety of Ml	High	High	Medium	High	Low
Maturity	High	Very High	Medium	Very High	Low

Machine Learning is very much in demand now a day. For making every machine self learning and based on the current results, the machines can predict further outcomes. Table 1 shows different parameters of various languages and tools .This languages and tools can be used for the implementation of various projects.

Machine Learning Analysis

Machine learning uses maths and statistics for many people. Statistics is a difficult subject used for companies to lie about how great their services. So why we need mathematics? The practice of Engineering is applied how to solve a problem. Machine learning is a process of transforming data into information. I this criteria the ML tool create a critical models. The developers can apply some different type of mechanism to improve the quality of modeling, testing, and refactoring.

The task in machine learning is a Regression Is the prediction of the data .The example of supervised learning is classification and regression. The opposite of supervised learning is a set of task known as unsupervised learning .There is nothing (targets) task given for the data. A clustering is a group of similar items. In unsupervised learning we want to find Statistical value to describe the data. It is called as density estimation.

Another task of unsupervised learning is either reducing the data from large task to a small number so we can properly visualize it in multiple dimensions. R is great language for machine learning for multiple reasons first it has a use of statistical; second it has a clear syntax. A multiple organizations use R, so there is large development and documentation.

Sensors and Data

We have fabulous amount of artificially created data from the internet but in current situation many nonhuman tools are available on online the technique behind that is not new, but linking them to internet is new.

Mobile phones or smart phones today ship with three axis magnetometer .The Smartphone also come with operating system .where you can execute your own programs; with a few lines of code you can get readings from the magnetometer 100 of times a second. Also, the phone already has it's a own communication system.

Flow chart for prediction system

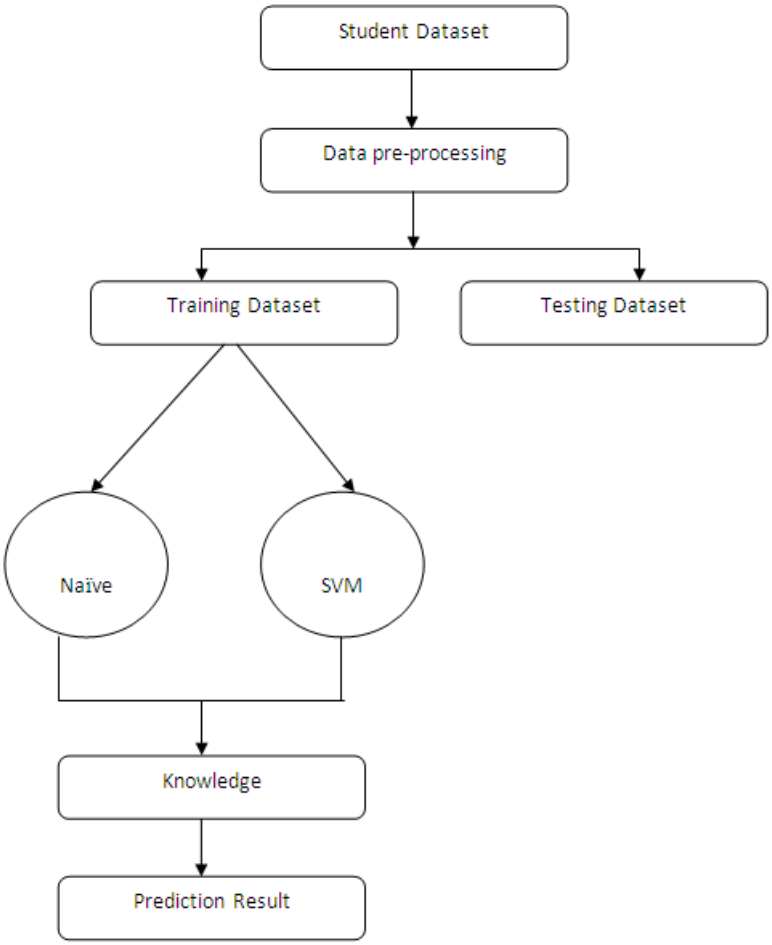

Fig. 1 - (a) Flowchart for Prediction System

2. Data Mining:

The data mining it is truly interdisciplinary concept. It can be defined as multiple ways. RDBMS is a set of multiple data and tables, which is assign an unequal name. Each table consists of set of columns and a large set of rows (tipples). Accented data model such as ER model (Entity Relationship), is constructed for Relational databases. Data have Quality if they satisfy the requirement of the intended used. There are many factors compressing data quality including completeness, timeline, accuses, interpretability and accuses.

Data cleaning routines work to remove the data by missing values, identifiers out layers. If user believes that the data are dirty, the data is not resalable of any data mining that has been applied to Data cleaning

3. Study of Algorithms

3.1. Support Vector Machine (SVM)

Support vector machine make good estimation of prediction for data points each are outside that training process.

Approach in SVM

1. Gather: any technique
2. Put together numeric values are wished
3. Three analyses it allows to visualize the keeping apart hyper traces
4. Train: the general public of the time can be spend right here. two parameters can be adjusted for the duration of this phase
5. Test: very simple calculation
6. You may use SVM in nearly any category problem. One issue be aware is that svm are binary classifiers. You will want to put in writing little more code to apply a svm on a hassle with more than two lessons.

There are two types of classifier of Support vector machine they are called machine because they develop a binary choice; they are choice machines. SVM have better generalization error; they establish good choice on what they learned. These advantages have made SVM algorithm famous.

The aim of SVM is to search for the ideal hyper plane separated by includes space and its help vector. Its acknowledgment needs processing help vector SV first, and after that registering ideal hyper plane (OHP). Since SV is the closest example purpose of OHP :(w·x) + b = 0, the separations between SV of a similar class and OHP are totally equivalent, yet not every one of the separations between SV of various classifications and OHP are equivalent. Consequently for certain m preparing tests:(,),(,),...,(,) 1 2 m x y x y x y , the way to registering a hyperplane is to process the parameters w and b. Since the hypothesis of help vector machine necessitates that hyper plane has littler arrangement blunder and better speculation capacity, it must meet the states of ideal hyper plane: y [ ] (w· x) + b ≥ 1, (I = 1,2,...,m) 2 min (w) w φ = keeping in mind the end goal to discover the ideal grouping hyper plane, as indicated by ideal hypothesis, the essential issue may be changed into process normal quadratic program with the help of Lagrange work: (,) 2 1 max ( ) 1 , 1 j I j I j m I j I m I W α ∑α α y K x = − s.t.0, 1 ∑ = I m I α y ≥ 0, α (I = 1,2,...,m) The way to registeringideal hyper plane is to figure α I when > 0 α I and: [](,(1))(,(1))21b0=Kw0x+Kw0x − α I , comparing to the most examples, is 0, and some αI(by and large a little part)isn't 0 (> 0 α I).Corresponding examples of those α I are bolster vectors and relating ideal hyper plane (i.e. characterization choice capacity) is: ( ) { (,) }0 0 f x sign I yiKxi x b I = ∑ − α > α The expectation of CEE desire is a multiple class grouping issue while the calculation is just a two-class order calculation.. More class arrangement issue and large illuminated by consolidating numerous two-class bolster vector machines. Mostly there's one to n and coordinated combinatorial example. For n-class arrangement issue, on the off chance that one-to-n combinatorial example is received, at that point n two-classification classifiers ought to be built up. In the event that balanced combinatorial example is received, at that point n (1)/2 − classifiers ought to be set up. Both the two techniques prompt that preparation and translating time increments quickly. It has been demonstrated that the limit of coordinated example is superior to anything one-to-n design. Thusly we, pick coordinated and utilize LIBSVM programming bundle to understand the foundation of n-class question classifier.

Multiclass SVM

- Multiclass SVM plans to allocate marks to cases by utilizing bolster vector machines, where the names are drawn from a limited arrangement of a few components.
- The predominant approach for doing as such is to diminish the single multiclass issue into numerous twofold arrangement issues. Basic techniques for such decrease include:
- Building twofold classifiers which recognize (I) between one of the marks and the rest (one-versus-all) or (ii) between each combine of classes (one- versus-one). Characterization of new occurrences for the one-versus-all case is finished by a victor takes-all methodology, in which the classifier with the most elevated yield work doles out the class (it is essential that the yield capacities be aligned to deliver similar scores). For the one-versus-one approach, arrangement is finished by a maximum wins voting procedure, in which each classifier allocates the case to one of the two classes, at that point the vote in favor of the relegated class is expanded by one vote, lastly the class with the most votes decides the occurrence grouping.
- Directed non-cyclic chart SVM (DAGSVM)
- Error-remedying yield codes

Crammer and Singer proposed a multiclass SVM strategy which throws the multiclass arrangement issue into a solitary enhancement issue, as opposed to disintegrating it into different parallel characterization problems. See additionally Lee, Lin and Wahba.

3.2. Bayes' Theorem

Naive Bayes' classifiers square measure a set of classification algorithms supported Bayes' Theorem. It is not one algorithmic rule however a group of algorithms wherever all of them share a typical principle, i.e. each combine of options being classified is freelance of every different.

To start with, allow us to contemplate a dataset. Consider a practical dataset that illustrate the climatic conditions for enjoying a game of golf. Given the climatic conditions, every tuple classifies the conditions as fit ("Yes") or unfit ("No") for playing golf. Following table represent our data

Table 2: Tabular Representation of dataset. [11]

Sr. No	Outlook	Temperature	Humidity	Windy	Play Golf
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

The above dataset is split into 2elements, namely, feature matrix and also the response vector. Feature matrix contains all the vectors (rows) of dataset within whichever vector consists of the worth of dependent options. In higher than dataset, options area unit 'Outlook', 'Temperature', Humidity and 'Windy'. Response vector contains the worth of sophistication variable (s or output) for every row of feature matrix. In higher than dataset, the category variable name is 'Play golf'.

Assumption:

The primitive Naive Bayes assumption is that every element makes an Individualistic Equivalent Contribution to the end result. With respect to our dataset, this idea is understood as: We consider that no try of options area unit dependent as an example, the temperature being 'Hot' has nothing to try to with the wetness or the outlook being 'Rainy' has no impact on the winds. Hence, the options area unit assumed to be freelance. Secondly, every feature is given constant weight (or importance). As an example, knowing solely temperature and wetness alone can't predict the end result accurately. None of the attributes is unsuitable and assumed to be conducive equally to the end result.

Note: In real-world things don't seem to be assumption appropriate which is created by Naive Bayes. In-fact, the independence assumption is rarely correct however typically works well in observe. It vital to know basic things regarding Bayes' theorem before going to understand formula of Naïve Bayes theorem.

Bayes' Theorem:
Bayes' algorithm searches the probability of an occasion happening given the probability of one more occasion which is just happened. Bayes' hypothesis is expressed numerically as the following condition:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \quad (1)$$

Where A and B area occasions and P (B)? Zero.

- Essentially, we tend to are attempt in to seek out chance of occasional, given the occasion B is valid. Occasion B is additionally termed as proof.
- P (A) is that the previous of A (the earlier chance, i.e. chance of event ahead proof is watched). The proof is Associate in Nursing attribute worth of Associate in Nursing unknown instance (here, it's occasion B)
- P (A|B) may be a posteriori chance of B, i.e. chance of occasion when proof is watched.

Now, within reference with above table of dataset, we are able to apply Bayes' theorem in the following route:

$$P(y \mid X) = \frac{P(X \mid y)P(y)}{P(X)} \quad (2)$$

Here, y is category variable and X may be a dependent upcoming vector (of size n) where:
X = (x1, x2, x3... xn)
Simply to clear, Associate in nursing example of an upcoming vector and examination category variable will be: (refer first row of dataset)
X= (Rainy, Hot, High, False)
Y = No
So essentially, P (X|y) here means, the chance of "Not taking part in golf" given that the atmospheric condition area unit "Rainy outlook", "Temperature is hot", "high stick" and "never wind".

Example:
Suppose we take two bags (bag c and bag d) with some BLUE pens and some RED pen in each bag. If it is given that blue pen is drawn then what is the probability that the pen is from bag c? Bayes theorem provides a formula to solve this condition based on reverse probability.
If,
B is the event when pen turns out to be BLUE
R is the event when pen turns out to be RED
C is the event when that pen drawn from bag c
D is the event when that pen drawn from bag d
Then, the probability of blue pen is drawn from bag c is

$$P(C / B) = \frac{P(B / C)}{P(B / C) + P(B / D)} \quad (3)$$

Where,
P (B/C) is the probability of finding a blue pen when a pen is drawn from bag c.
P (B/D) is the probability of finding a blue pen when a pen is drawn from bag d.

3.3. Data-intensive computing

Information escalated figuring is a class of parallel registering applications which utilize an information parallel way to deal with process expansive volumes of information ordinarily terabytes or petabytes in estimate and normally alluded to as large information. Figuring applications which give the majority of their execution time to computational necessities are esteemed register serious, though processing applications which require vast volumes of information and commit the majority of their handling time to I/O and control of information are considered information concentrated.

3.3.1. Parallel processing:

Parallel handling approaches is for the most part isolated into two kinds:
Register serious
Data-serious.
Register serious is utilized to depict application programs that are figure bound.
Such applications commit a large portion of their execution time to computational necessities rather than I/O, and normally require little volumes of information. Parallel handling of register serious applications regularly

includes parallelizing singular calculations inside an application procedure, and breaking down the general application process into discrete undertakings, which would then be able to be executed in parallel on a fitting figuring stage to accomplish by and large higher execution than serial preparing. In register concentrated applications, numerous tasks are performed at the same time, with every activity tending to a specific piece of the issue. This is regularly alluded to as errand parallelism

4. Compute-Intensive

Process concentrated is a term that applies to any PC application that requests a considerable measure of calculation, for example, meteorology programs and other logical applications. A comparative however unmistakable term, PC concentrated, eludes to applications that require a considerable measure of PCs, for example, lattice figuring. The two sorts of uses are not really fundamentally unrelated: a few applications are both figure and PC concentrated.

5. Comparison between Existing System and Proposed system

Table 3: Comparison between Existing System and Proposed system

Contents	Existing	Proposed
Human Resource	More	Less
Paper work	More	Less
Time	Time Consuming	Less time required
Cost	Expensive	Less Expensive
Complexity	complicated	Easy

Implementation strategy for proposed system

The system is prediction model for college prediction purpose which uses R language, R studio and a data mining technique for the prediction. The dataset has been collected over the particular institute from the past dataset. The dataset is being pre-processed to make the dataset valid. The pre-processed data helps to enhance the performance of the system and also improves the accuracy of the system.

Acknowledgment: The authors would like to thank Prof. Mahendra Patil for his helpful discussion related to this paper.

III. Conclusion

The system uses a hybrid model for predicting admission of students in the institute's .The use of SVM Linear gives a more accurate result than the earlier systems as SVM Linear is faster and memory efficient as compared to other algorithms. The proposed system has successfully showcased the ability to get the admission in engineering based on their HSC and CET marks. The brief demonstration about the system is mentioned to access the information based on database. It gives the better description about the action taken on the previous data. Nevertheless, there are many improvised systems that show the betterment of this system. The system makes use of more number of attributes for more accuracy.

References

[1]. Z. Wang and Y. Shi, "Prediction of the Admission Lines of College Entrance Examination based on machine learning", IEEE, 2016, s7, 10.
[2]. C.Chrysostomou, H. Partaourides, H.Seker, "Prediction of Influenza A virus infections in humans using an Artificial Neural Network learning approach", IEEE-EMBC, 2017, s12.
[3]. S. Mishra, S. Sahoo, S. Mishra and S. Satapathy, "A Quality Based Automated Admission System for Educational Domain", IEEE, 2016, s10.
[4]. Xiaoyu Sun, Xia Yu and Honghai Wang, "Glucose prediction for type 1 diabetes using KLMS algorithm" IEEE, 2016, s 7, 10.
[5]. R. Jia, R. Li, M. Yu and S. Wang, "E- commerce purchase prediction approach by user behavior data", CITS, 2017, s12.
[6]. D. Vaghelaand P. Sharma, "Students' Admission Prediction using GRBST with Distributed Data Mining", CAE, 2015, s 9, 12.
[7]. C. Mirji, V. Deshpande, S. WalunjandIOSR- JCE, 2014, s 11.
[8]. J.Bibodi, A.Vadodaria, A.Rawat and J.Patel,"Admission Prediction System Using Machine Learning", IJSRC, 2012, s 10.
[9]. R. Dong, H. Wang and Z. Yu, "The Module of Prediction of College Entrance Examination Aspiration", FSKD, 2012, s 10.
[10]. Ragab, A. Mashat and A. Khedra, "HRSPCA: Hybrid Recommender System for Predicting college Admission", IEEE, 2012, S8, 12.
[11]. https://www.geeksforgeeks.org/naive-bayes-classifiers/