ENGINEERING & TECHNOLOGY ADMISSION ANALYSIS AND PREDICTION

Mr. Sachin Bhoite, Ph. D. Research student, Bharati Vidyapeeth Deemed University, Pune

Prof. Dr. Ajit More, Director MCA Program, IMED, Bharati Vidyapeeth Deemed University, Pune

Email: sachin.bhoite@mitwpu.edu.in

Abstract*: A Great career without a Great Education is just a DREAM. While we talk about career – a person's degree, specialization, College/University and the knowledge that he possesses – are the key factors. In India the educational pattern is 10+2+3+2 or 10+2+4+2 or 10+2+5.5 & career related decisions are discussed after 10th standard and mostly concluded after 12th. As soon as a student completes his/her Higher Secondary Schooling, the first goal of any student is to get into an appropriate College/University for appropriate course/program so that he can get a better education, guidance & placement for his future.*

To build predictive model we used Logistic Regression, K Nearest Neighbors', Decision Tree Classifier, Random Forest Classifier, Naive Bayes & Support Vector Machine classifiers then compare the results of cross-validation with & without feature engineering and also compare the probabilities of getting admission to a college. The performance of various classifiers is described in this paper. It is found that Random Forest & Decision tree classifiers give better accuracy.

Keywords*: Logistic Regression, K Nearest Neighbors', Decision Tree Classifier, Random Forest Classifier, Naive Bayes & Support Vector Machine*

1. INTRODUCTION

Classification is one of the most frequently studied problems by DM and machine learning (ML) researchers [1]. It consists of predicting the value of a (categorical) attribute (the class) based on the values of other attributes (the predicting attributes) [4]. The aim of this paper is to determine the factors estimating & guiding the students to select engineering college for their first year admission. Most of students & parents are spending unnecessary efforts, time & money on selecting right engineering college for first year admission. Sometimes the students who are seeking admission is not eligible to take admission into engineering program based on their past academic record. Also sometimes the students are seeking the admission to the college for she or he are not eligible as per the merit of that college. So here researcher has built predictive model to guide the students about their admissibility in the desired college & also suggest the college where they will get the admission. So to achieve this objective we may include machine learning capabilities that allow to improve their performance based on experience, just as humans do. [3] As right College plays very vital role form the students' placement and career point of view the researcher has implemented various algorithms to achieve this objective.

Basically this model will help to save time, money & mainly confusion of predicting right alternatives for engineering education after 12th. Which will also help in arranging or planning of future expenses.

As our objective is to predict whether a student will get an admission or not at the desired Institute & also the rank wise list of possible colleges where they will get admission. So it's majorly classification problem. Therefore we used Logistic Regression, K Nearest Neighbors', Decision Tree Classifier, Random Forest Classifier, Naive Bayes & Support Vector Machine Supervised Machine Learning Algorithms.

2. RELATED WORKS

The researchers have studied various related national & international research papers, thesis to understand objectives, type of algorithms used, datasets, data preprocessing methods, features selection methods, etc.

Ahmad Slim, Don Hush, Tushar Ojah, Terry Babbitt (2018) used logistic regression (LR), support vector machines (SVMs) and semi-supervised probability machine learning methods for Predicting Student Enrollment Based on Student and College Characteristics. The LR and the SVMs methods predict the enrollment of applicants at an individual level whereas the semi-supervised probability method does that at a cohort level. A small set of factors related to student

and college characteristics are highly correlated to the applicant decision of enrollment. This outcome is supported by the relatively high prediction accuracy of the proposed methods [3].

Min NIE1, Lei YANG1, Jun SUN1, Han SU1, Hu XIA1, Defu LIAN1, Kai YAN (2017) propose system for advanced forecasting of career choices for college students based on campus big data. They used logistic regression (LR), SVM, Random Forest (RF), and Decision Tree (DT). Random Forest performs comparatively better than the others. Based on four types of behavior features, evaluated the effectiveness of framework and found that the extracted professional skills, behavior regularity, and economic status were significantly correlated with career choices [4].

K. Sripath Roy 1, K.Roopkanth 2*, V.Uday Teja 3, V.Bhavana4, J.Priyanka 5 (2018) To predict students career advanced machine learning algorithms like SVM, Random Forest decision tree, OneHot encoding, XG boost are used. This research will help Computer Science students as well as recruiter. Out of all SVM gave more accuracy with 90.3 percent and then the XG Boost with 88.33 percent accuracy[5].

Prof. M. L. Waghmode, Dr. P.P. Jamsandekar (January 2016) designed framework for the expert system useful for career selection on the factors considered are Ability: Skills to do something, Attitude, Personality & Interest. Machine learning algorithms ID3, PRISM and PART gives 100% accuracy in classification along with rules[6].

Nikita Gorad1, Ishani Zalte2, Aishwarya Nandi3, Deepali Nayak4 (April 2017): According to the survey conducted by the Council of Scientific and Industrial Research's (CSIR), about 40% of students are confused about their career options. This system is a web application that would help students studying in high schools to select a course for their career. The system would recommend the student, a career option based on their personality trait, interest and their capacity to take up the course. The prediction is done using one of the Decision tree algorithms which is C5.0 package in R. Using this system, it recommends a particular course along with the list of colleges providing those courses [7].

Mr. Shashikant Pradip Borgavakar Mr. Amit Shrivastava (May - 2017) The k-means clustering algorithm is used to classify student's grade on basis of class test, mid test and final test into three categories: "High" "Medium" and "Low". The students with low class are identify as the weak students and they can guide by the teachers before final examination [8].

V. Sai kuswanth, Gogineni Krishna Chaitanya, B.Sekhar Babu, Uppuluri Lakshmi Soundharya (March 2019): In this expert system artificial neural network is applied for student career assessment on the basis of student marks in mathematics and physics and some questions are given to students and answering these questions ensures the student's capabilities and analyses their

capabilities whether they are perfect for job which they are going to do in future.[9]

Dr. A. Padmapriya* (November 2012) Data mining algorithm Decision Tree Induction is best when compared to Naive Bayesian Classifier according to Classification Accuracy reMisclassification Rate, Speed & Size on students Personal Data, Pre college data & Under Graduation Data to predict Higher Education Admissibility [10].

Jay Bibodi (2019) has done work on predicting the university and students applying to explicit universities. He came across some issues like noisy data, unformatted text but after cleaning the data, they proceeded to 'model selection' with some important features. "University Selection Model" – A Classification problem with apriori probability output. They found out just two universities giving a higher probability of output. "Student Selection Model" – Classification using supervised learning like Linear and kernel, Decision Tree and Random Forest. Random Forest provided better accuracy than other algorithms i.e. 90% accuracy [11].

Himanshu Sonawane (2019) who has explored 'Student Admission Predictor'. It is a system built to help students who are studying abroad. This system helps students find the best foreign universities/colleges based on their performance in GRE, IELTS, 12th, Graduation Marks, Student Statement of purpose, Letter of Recommendation, etc. Based on this information, it recommends the best-suited university/college. They have used three algorithms: KNN (76% Accuracy), Decision Tree (80% Accuracy) and Logistic Regression (68% Accuracy). In the case of a decision tree, accuracy was nearly the same for both pieces of training as well as testing datasets [12].

H. Sabnani, M. More, P. Kudale, S. Janrao, (2018) proposed Prediction of Student Enrolment Using Data Mining Techniques. They have used the Apriori technique to analyze the behavior of students who are seeking admission to a particular college. They have also used the Naïve Bayes algorithm which will help students to choose the course and help them in the admission procedure. In their project, they were conducting a test for students who were seeking admissions and then based on their performance, they were suggesting students a course branch using Naïve Bayes Algorithm[13]

Bhavya Ghai (2018) has presented study on Analysis & Prediction of American Graduate Admissions Process for helping students in suggesting them best-suited colleges in the USA based on his/her profile. He has collected the data from online sources which was reported by students. He has used 5-6 algorithms for his project. Naïve Bayes was one of them which gave the highest accuracy among all of them. He has predicted students' chances (probabilities) of getting admission in 5 different universities in the USA [14].

Dineshkumar B Vaghela, Priyanka Sharma (2015) has presented Students' Admission Prediction using GRBST with Distributed Data Mining. They have used the Global Rule Binary Search Tree (GRBST). While searching, they identified some problems like maintaining a single database for

all the colleges were difficult. This paper has two phases i.e. training phase and testing phase. In the training phase, the J48 algorithm was used for all local sites. In the testing phase, Users can interact with the system with the help of the application layer. They have used consolidation techniques in two ways i.e. using If…Then… rules format and Decision Table. They have also used binary search tree construction. After applying this technique, they have found the time complexity of generating the Binary Search Tree from the Decision table is very less and also this BST has efficient time complexity to predict the result. They conclude that data mining techniques can be useful in deriving patterns to improve the education system [15].

3. RESEARCH METHODOLOGY

To achieve objective step-by-step are involved to make the data steady, fitting it into models and finding out suitable algorithms of machine learning for building a predictive model.

3.1 RESEARCH DESIGN

The below architecture follows three different steps. First step is all about dataset & preparing input variables & output variables ready for predictive modeling. Step-II & Step-III worked together in synchronized manner. The important part of this predictive modelling is comparison & estimation of accuracy of the model. So for this purpose K-fold cross validation method is used. Then the train & test dataset are being provided to all the models for fitting & predict purpose respectively. Finally again best accuracy model is selected for predicting value of dependent variable of single record. These steps are shown in Figure 1.

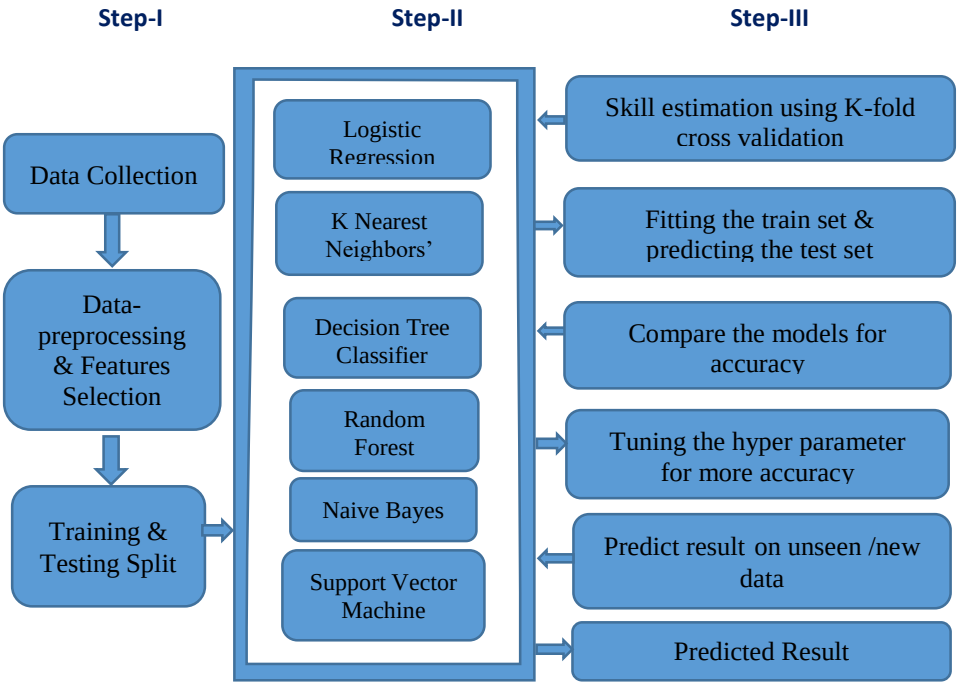
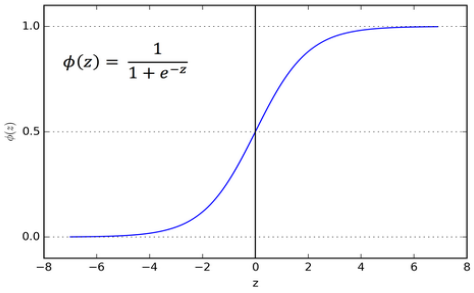


Figure 1: Architecture

3.2. METHODOLOGY

3.2.1 Logistic Regression

Logistic Regression is used for supervised classification problem. The regression model used to predict probability of a given data. The Logistic function is also called as sigmoid function (which range from 0 to 1). Sigmoid function is use to map predictions to probabilities.



The logistic regression equation is

Y = e^(B0 + B1X1 + . . . + BnXn) / 1 + e^(B0 + B1X1 + . . . + BnXn)

Where Y is dependent variable (predicted output), B0 is bias term, X is independent variable, n = number of observations, B1 is coefficient to input value X. Logistic regression is a linear regression but logistic regression uses more complex cost function compare to linear regression.

The sigmoid function is given as Linear Equation and Sigmoid Function

$$z = \theta_0 + \theta_1 \cdot x_1 + \theta \cdot x_2 + \cdots \quad g(x) = \frac{1}{1+e^{-x}} \qquad h = g(z) = \frac{1}{1+e^{-z}}$$ [20].

3.2.2. SUPPORT VECTOR CLASSIFIER (SVC):

This algorithm is used for the classification problem. The main objective of SVC is to fit the data you provide, returning a “best fit” hyperplane that divides or categorizes your data. From there, when obtaining the hyperplane, you'll then feed some options to your category to examine what the "predicted" class is.

3.2.3. DECISION TREE

A decision tree is non-parametric supervised learning. It is used for both classification and regression problems. It is a flowchart-like structure in which each internal node represents a “test” on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label. The path between root and leaf represents classification rules. It creates a comprehensive analysis along with each branch and identifies decision nodes that need further analysis

3.2.4. RANDOM FOREST

Random Forest is a meta estimator that uses the number of decision trees to fit the various subsamples drawn from the original dataset. We can also draw the data with replacement as per the requirements.

3.2.5 K-NEAREST NEIGHBORS (KNN)

K-Nearest Neighbors (KNN) is a supervised learning algorithm that is used to solve regression as well as classification problems. Where 'K' is the number of nearest neighbors around the query. It is simple to implement, easy to understand and it is a lazy algorithm. The lazy algorithm means it does not need any training data points for a model generation [17]. All the training data is used in the testing phase. This makes the training faster and the testing phase slower and costlier. By costly testing phase we mean it requires more time and more memory.

3.2.6. NAÏVE BAYES:

A Naive Bayes Classifier is a supervised machine- learning algorithm that uses Bayes' Theorem, in which the features are statistically independent. It specifies multiple uses of probability theories and statistics. By simple machine learning problem, where we need to teach our model from a given set of attributes (in training examples) and then form a hypothesis or a relation to a response variable. Then we tend to use this to predict a response, given attributes of a replacement instance.

3.3 ABOUT THE DATASET

To achieve all the objectives as per the framework to guide aspirant Engineering graduates & their parents we need past admission data of multiple Engineering colleges to work on. So we have collected various non-aided but affiliated to Savitribai Phule Pune University's engineering colleges admission record of academic year 2015-16 for the study

In the dataset there are no records which shows student had applied but not admitted due to their poor merit score. As only these records are available with Colleges & Joint Director office, Pune. This Dataset has various attributes, which are: 'Main Serial No.', 'Sr. No.', 'College Name', 'College Code', 'Merit No', 'Merit Marks', 'Candidate Name', 'Gender', 'Candidate Type', 'Category', 'Home University', 'PH Type', 'Defence Type', 'HSC Eligibility', 'Seat Type', 'Fees Paid', 'CAP Round', 'Admitted/Uploaded Late', 'BRANCH' and 'NATIONALITY'.

3.4 FEATURE ENGINEERING

Data pre-processing is an important task in machine learning. It converts raw data into clean data. Following are techniques, we have applied on data: -

Missing Values – Missing Value are those values that failed to load information or the data itself was corrupted. There are different techniques to handle missing values. One of which we have applied is deleting rows because some of the rows were blank and they may mislead the

classification.

In the first dataset missing values are as follows.

College Name	0
College Code	0
Merit No	0
Merit Marks	0
Candidate Name	0
Gender	0
Candidate Type	0
Category	458
Home University	3
PH Type	14058
Defence Type	13776
HSC Eligibility	0
Seat Type	0
Admitted/Uploaded Late	0
BRANCH	0
NATIONALITY	0

Out of 20 columns only Category, Home University, PH Type and Defence Type columns are having missing values. As per actual dataset PH Type & Defence Type registered candidates are less than 5%. So remaining 95% cells are empty for these two columns, which shows missing values. To fill remaining cells of the columns we apply new PH & Defence Category which is 'NA'. As per dataset category column has approximately less than 5% missing values. So to fill it we fill that cells with "Open" category.

● Handling Categorical Data (Label Encoder) – This is one of the most frequently used techniques for the categorical variable. Label encoder converts labels into a numeric format so that the machine can recognize it. In our data, there are many attributes which are categorical variable like Gender, Candidate Type, Category, Home University, Branch, Nationality, College Name and CAP Round, PH Type, Defence Type, BRANCH, NATIONALITY.

● Change in data type – Some attributes didn't include proper input. For example, the "Home University" attribute included values like 'PUNE', 'PUME', 'PUNE ' which all meant the same University. For that purpose, we needed to change such values into a single format.

We don't need to change any data type except label encoding of object data type to int32.

● Drop Columns – As per domain knowledge, we initially removed some columns which were not needed in our model which are: 'Main Serial No.','Sr. No.', 'Candidate Name', 'Seat Type', 'Fees Paid' and 'Admitted/Uploaded Late'.

4. FEATURE SELECTION

Every time domain experts of the problem may not be available to decide independent

features to predict the category of the target feature. Hence before fitting model we must make sure that all the features that we have selected are contributing to the model properly and weights assigned to it are good enough so that our model gives satisfactory accuracy. For that, we have used 3 feature selection techniques: Univariate Selection, Recursive Features Importance, & Lasso - feature selection.

For Univariate Selection, following feature importance is generate.

Feat_names	**F_Scores**
HSC Eligibilty	200.487740
Merit Marks	93.718291
BRANCH	77.121620
Candidate Type	22.456847
Home University	5.294923
Defence Type	4.073951
Category	2.205066
PH Type	1.327218

As per this technique H.S.C. Eligibility & Merit Marks are important features. While using Recursive Feature Importance method following features are selected and remaining are rejected.

Selected Features: ['Candidate Type', 'Category', 'PH Type', 'Defense Type', 'HSC Eligibility', 'BRANCH']

As per Lasso feature selection method importance is given to following features.

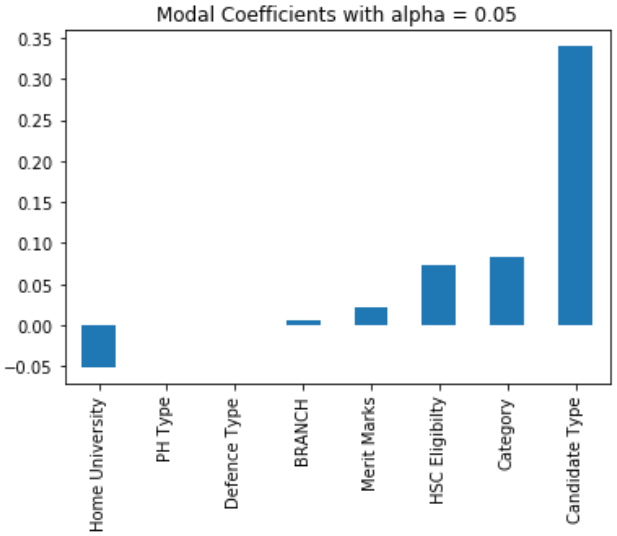


Figure 2: Feature Selection using LASSO

As per above figure top features sorted scores:

Features	Lasso Score
Candidate Type	0.339628
Category	0.0823817
HSC Eligibility	0.0733778
Home University	0.0512474
Merit Marks	0.0221583
BRANCH	0.00598959

As we can see all above Feature Selection methods are giving importance to different features. So, in this case, our domain knowledge is also helpful to make decisions for this type of situation.

5. EXPLORATORY DATA ANALYSIS

As we saw in feature selection, some features which seemed not so important were contributing to our model. So, to understand those features, we need to do exploratory analysis on this data. We did exploratory analysis on a few features by grouping and plotting it on graphs.

College Name and Gender Column: By grouping the gender and plotting the admissions in different colleges as per their gender, we identified some relations between the student's admission and his or her gender. As shown in Figure 3 –except BVCE, Katraj College all the colleges shows number of boys are more than number of girls in count (almost double). As BVCE is women's college, all candidate registered are girls. But it's not impacting on admissibility of the students. As per all three feature selection techniques (top f_classifier Recursive Features Importance, & Lasso - feature selection) Gender feature is rejected.

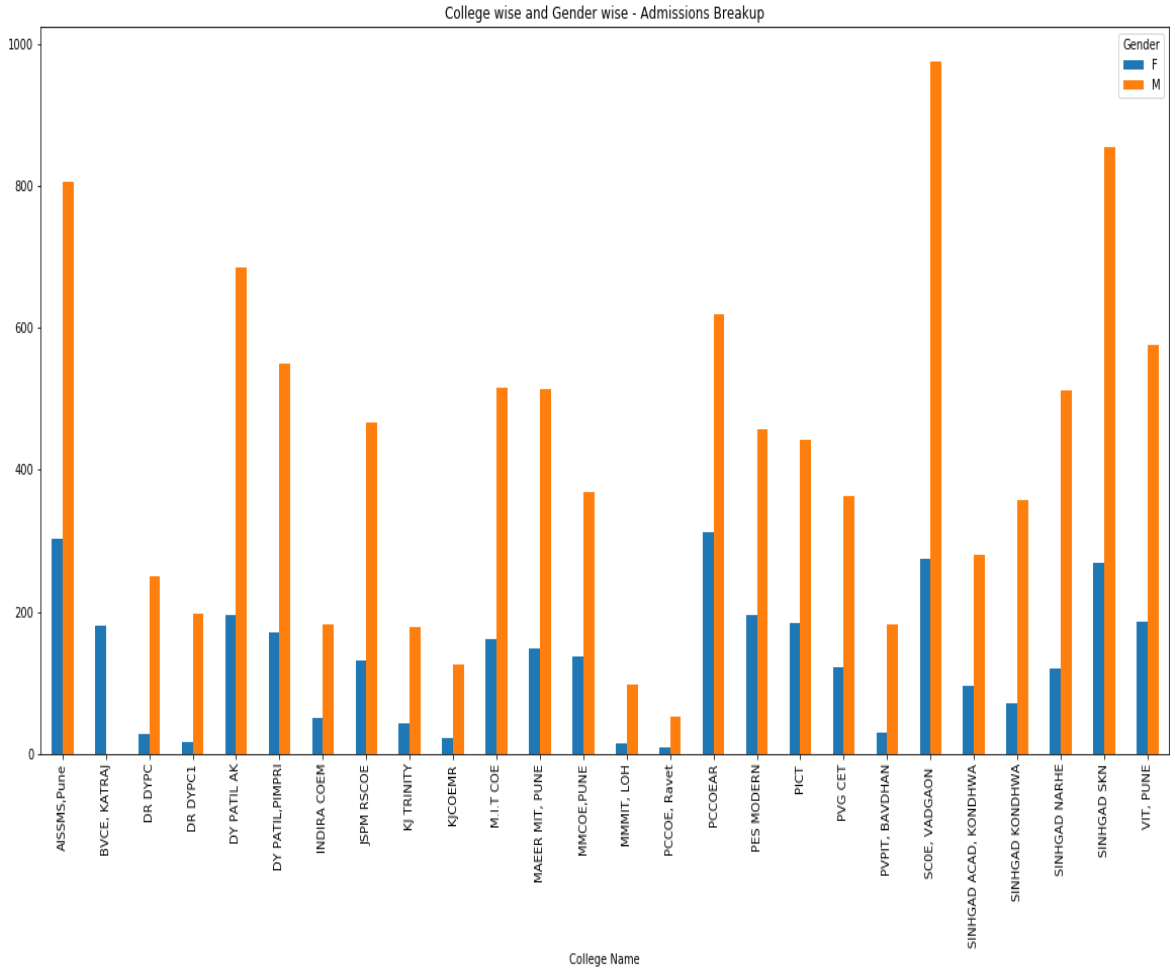


Figure 3: College Name and Gender count plot

College Name & Category Column: By grouping the category the students who got admissions

with respect to their categories is shown in Figure 4 –. The percentage of students was matching to reservation criteria as per reservation laid down by Government of Maharashtra & DTE. This shows that the Category column is contributing to our model. As per all three feature selection techniques (Univariate Selection, Recursive Features Importance, & Lasso - feature selection) Category feature is selected.

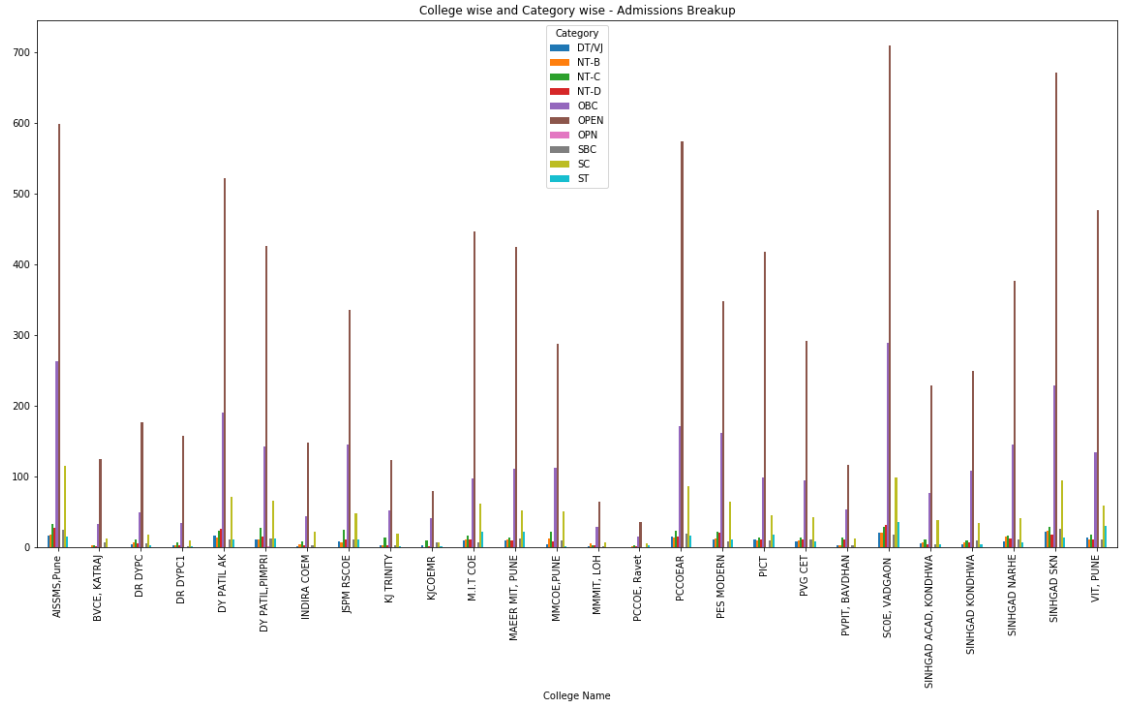


Figure 4: College Name and Category count plot

College Name & Branch Column: The branch is one of the important feature from admission point of view. You can see total 20 branched are available out of which 'COMPUTER', 'MECHANICAL' & ‘CIVIL’ branches of all colleges show maximum admissions. As per all three feature selection techniques (Univariate Selection, Recursive Features Importance, & Lasso - feature selection) Branch feture is selected.

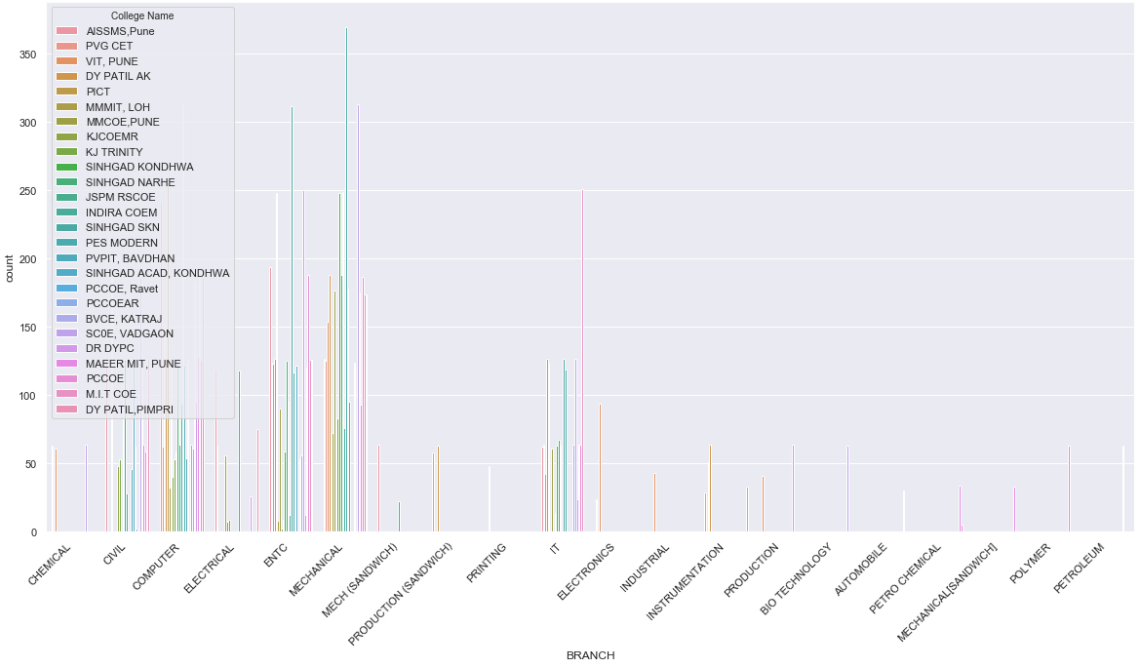


Figure 5: College Name and BRANCH count plot

College Name & Home University: As per graph students from 11 different Universities are registered, out of which as per graph count of PUNE University is comparatively very high, there after count from NAGPUR, North Maharashtra and BAMU. As per all three feature selection techniques (Univariate Selection, Recursive Features Importance, & Lasso - feature selection) Home University feature is selected.

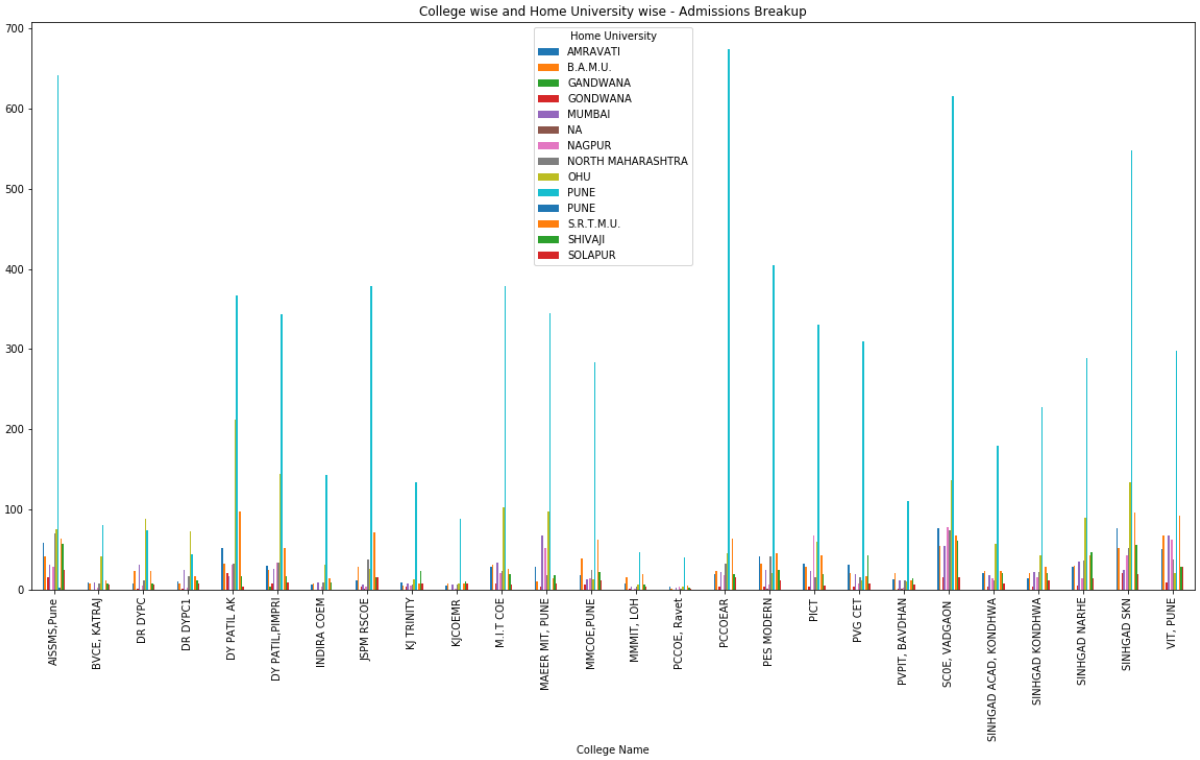


Figure 6: College Name and Home University count plot

College Name & Candidate Type Column: There are 13 types of categories of the candidate type, Out of which as per graph TYPE A candidate registered are maximum as compared to other candidates As per all three feature selection techniques (Univariate Selection, Recursive Features Importance, & Lasso - feature selection) Home University feature is selected.

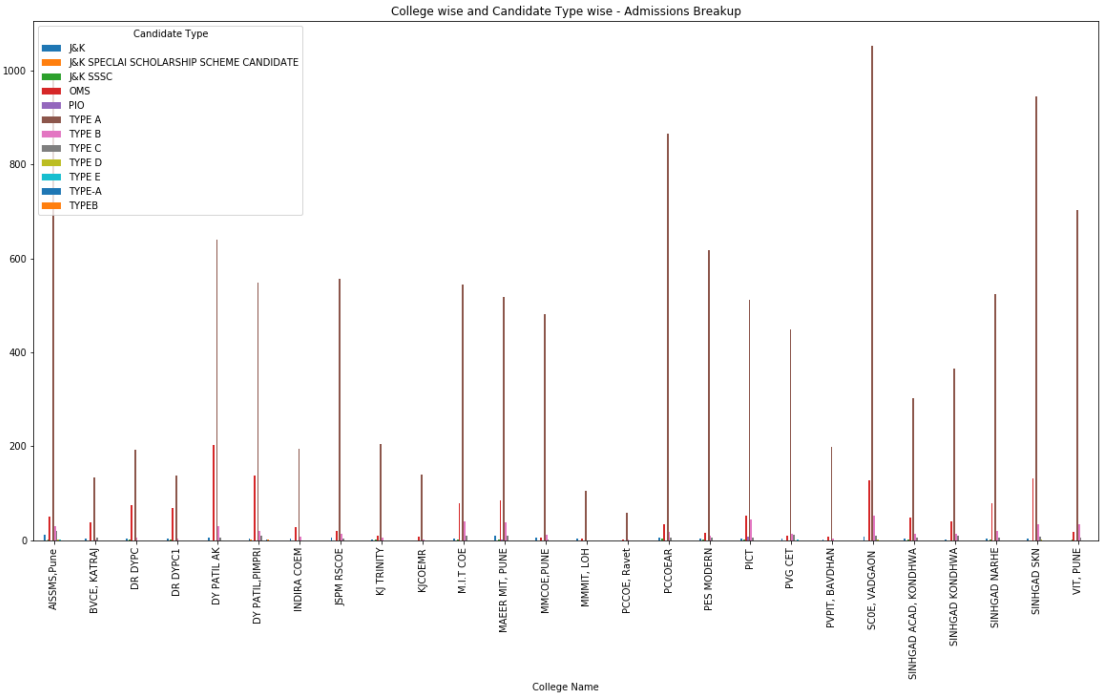


Figure 7: College Name and Candidate Type count plot

College Name & HSC Eligibility Column:

As per following box plot HSC merit of PICT, Pune College is high compared to other colleges as students admitted to this college are secured high marks in their H.S.C. As per all three feature selection techniques (Univariate Selection, Recursive Features Importance, & Lasso - feature selection) HSC Eligibility feature is selected.

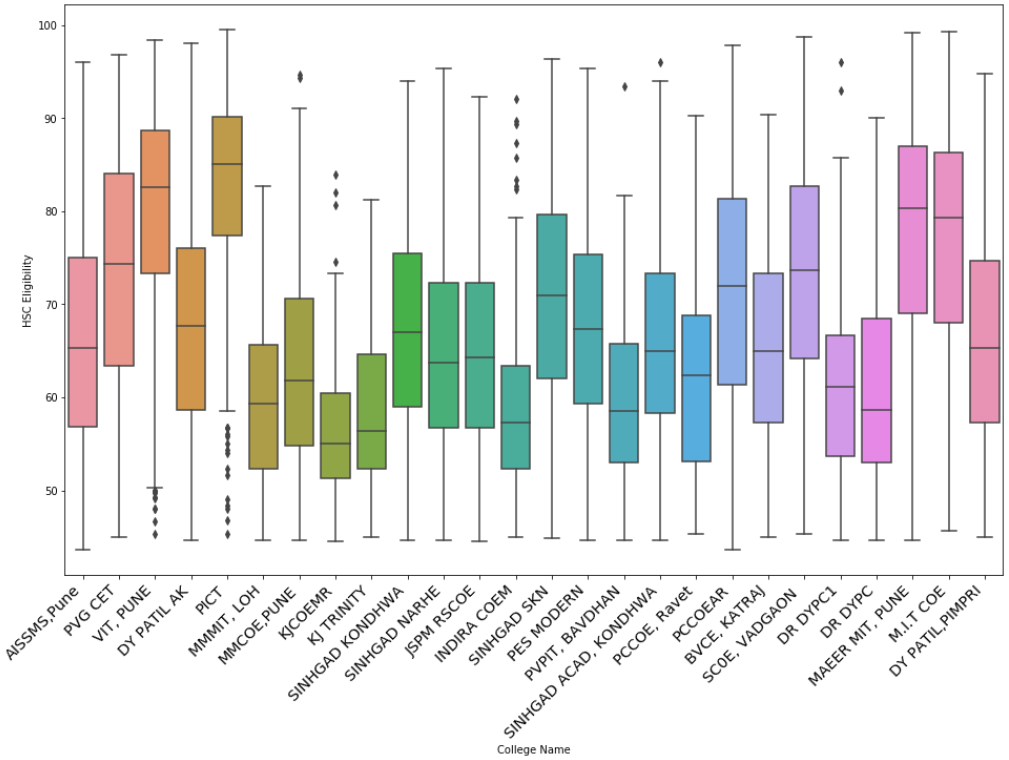


Figure 8: College Name and HSC Eligibility box plot

College Name and Merit Marks Column: In the following graph also shows that Merit Marks (JEE Mains Score+ HSC Marks) of PICT Pune college is high as compared to other colleges. As per all three feature selection techniques (Univariate Selection, Recursive Features Importance, & Lasso - feature selection) Merit Marks feture is important.

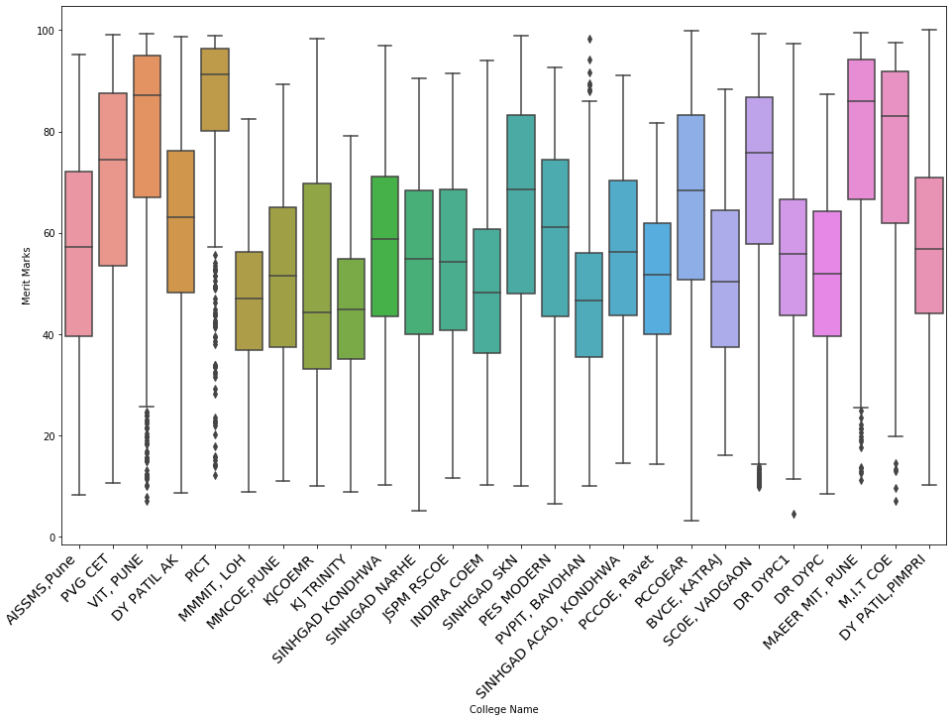


Figure 9: College Name and Merit Marks box plot

In this way we have checked importance of all the features and selected it with the help of feature selection method along with EDA. Also EDA plays very important role in analysis of feature wise data & their correlation with each other.

6. MODEL IMPLEMENTATION AND RESULTS

After cleaning all the data, removing all the noise, selecting relevant features and encoded it into machine learning form, the next step is build a predictive model by applying various ML techniques to find out the best model which gives us more accuracy for train and test both. But before that, we must split our data into 2 parts.

6.1 TRAIN-TEST SPLIT

The training data set is used to create the model while testing the data set is used to check the performance. Training data's output is available to model while test data is unseen data. So, in our data, we have split data into 70% for training data and 30% for testing data because it makes the classification model better. While the test data makes the error estimate more accurate [19].

6.2. RESULT COMPARISON (RAW DATA VS FEATURED DATA)

With cross validation method we have used 6 standard algorithms for prediction of college which includes Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Gaussian Naïve Bayes and K Neighbors. All the chosen appropriate algorithms run through test harness of number of splits 10 and performance major accuracy. We have shown accuracy results of it before feature engineering and after feature engineering, which are as follows.

The cross validation accuracy results of College prediction model:

Sr. No.	Name of Algorithm	Accuracy
1	Logistic Regression	0.177343
2	Support Vector Machine(SVC)	0.168846
3	Decision Tree Classifier	1.000000
4	Random Forest Classifier	0.998426
5	Gaussian NB	0.548710
6	K Neighbors Classifier	0.357517

Table 1: Accuracy score with cross validation for college prediction model before feature engineering.

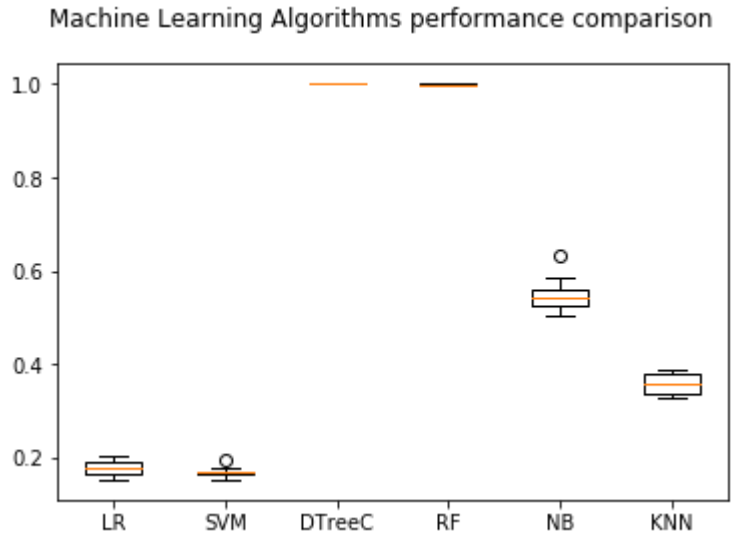


Fig. 10: Accuracy score comparison of college prediction model before feature engineering.

Sr. No.	Name of Algorithm	Accuracy
1	Logistic Regression	0.478524
2	Support Vector Machine(SVC)	0.498821
3	Decision Tree Classifier	0.985682
4	Random Forest Classifier	0.971050
5	Gaussian NB	0.950275
6	K Neighbors Classifier	0.601098

Table 2: Accuracy score with cross validation for college prediction model after feature engineering.

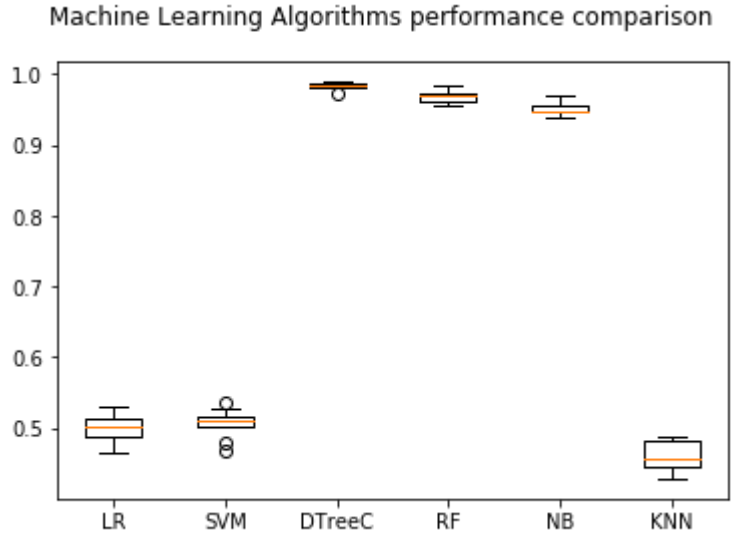


Fig. 11: Accuracy score comparison of college prediction model after feature engineering.

After observing accuracy results of the model before feature engineering and after feature engineering we can simply understand importance of feature engineering, while building predictive model using Machine Learning techniques. It has been observed that accuracy of classifiers Logistic Regression, Support Vector Machine, Gaussian NB and KNN are improved more after feature engineering and accuracy of classifier DTree & Random Forest are seen more practical which were initially turned towards the over fitting before feature engineering. After finalizing classifier of the model sometimes we required tuning the model for better accuracy. Tuning is the method for increasing a model's performance without over fitting the data or making the variance too high. Hyper parameters disagree from other model parameters therein they're not learned by the model automatically through training ways [19]. Further we have taken single record to classify the college & tested with various classifiers. For this single record we have again build all six predictive model and tested for different input variable. So we found every time that out all six models Decision Tree classifier & Random Forest always give great accuracy.

7. CONCLUSION

In this paper, the researchers describes the Architecture for hassle-free College prediction, compare cross-validation techniques for accuracy as performance major before and after feature engineering and finally concluded the prediction of College for career after 12th in the engineering and technology. For this study 8 input features are selected out of 20 features, which are 'Merit Marks', 'Candidate Type', 'Category', 'Home University', 'PH Type', 'Defense Type', 'HSC Eligibility', and 'BRANCH'. These features are very important according to Univariate Selection, Recursive Features Importance, & Lasso feature selection methods and massive Exploratory Data Analysis used by checking and plotting correlation between each input feature with target feature. In this study target is categorical in nature. So we used Logistic regression, K-NN, Decision Tree classifier, Random Forest, Naïve Bays, Support Vector Machine 6 supervised machine learning algorithm. We found out of six algorithms Decision Tree classifier & Random Forest gives good & approximately same accuracy which is 93. Also we realize that feature engineering is very essential part while implementing & building predictive models using machine learning techniques. It has been observed that results has been more improved after feature engineering. In future we would like to consider students choice about the college as an input feature and mix i.e. admitted and rejected student’s data to get better accuracy.

REFERENCES

1. Kabakchieva, D., Stefanova, K., Kisimov, V. (2011). Analyzing University Data for Determining Student Profiles and Predicting Performance. Conference Proceedings of the 4th International Conference on Educational Data Mining (EDM 2011), 6-8 July 2011, Eindhoven, The Netherlands, pp.347-348.
2. Ma, Y., Liu, B., Wong, C. K., Yu, P. S., Lee, S. M. (2000). Targeting the right students using data mining. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, Boston, pp 457-464.
3. Ahmad Slim, Don Hush, Tushar Ojah, Terry Babbitt (2018) Predicting Student Enrollment Based On Student And College Characteristics, International Educational Data Mining Society, Paper presented at the International Conference on Educational Data Mining (EDM) (11th, Raleigh, NC, Jul 16-20, 2018)
4. Min NIE1, Lei YANG1, Jun SUN1, Han SU1, Hu XIA1, Defu LIAN1, Kai YAN (2017) Advanced Forecasting Of Career Choices For College Students Based On Campus Big Data _ Higher Education Press and Springer-Verlag Berlin Heidelberg 2017
5. K. Sripath Roy 1, K.Roopkanth 2*, V.Uday Teja 3, V.Bhavana4, J.Priyanka 5 (2018) Student Career Prediction Using Advanced Machine Learning Techniques International Machine Learning Algorithm, Journal of Engineering & Technology, (2.20) 26-29
6. Prof. M. L. Waghmode, Dr. P.P. Jamsandekar (January 2016) Expert System For Career Selection: A Classifier Model International Journal of Advance Research in Computer Science and Management Studies ISSN: 2321-7782
7. Nikita Gorad1, Ishani Zalte2, Aishwarya Nandi3, Deepali Nayak4 (April 2017) Career Counselling Using Data Mining International Journal of Innovative Research in Computer and Communication Engineering ISSN(Online): 2320-9801 ISSN (Print): 2320-9798
8. Mr. Shashikant Pradip Borgavakar Mr. Amit Shrivastava (May - 2017) Evaluating Student's Performance Using K-Means Clustering International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181
9. V Sai kuswanth, Gogineni Krishna Chaitanya, B.Sekhar Babu, Uppuluri Lakshmi Soundharya (March 2019) Machine Learning Approach For Students Career Assessment In The Modern World International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6S, March 2019.
10. Dr. A. Padmapriya* (November 2012) Prediction Of Higher Education Admissibility Using Classification Algorithms International Journal of Advanced Research in Computer

Science and Software Engineering ISSN: 2277 128X

11. Bibodi, J., Vadodaria, A., Rawat, A. and Patel, J. (n.d.). "Admission Prediction System Using Machine Learning". California State University, Sacramento.
12. Himanshu Sonawane, Mr. Pierpaolo Dondio. "Student Admission Predictor". School of Computing, National College of Ireland.
13. H. Sabnani, M. More, P. Kudale, S. Janrao, "Prediction of Student Enrolment Using Data Mining Techniques", International Research Journal of Engineering and Technology (IRJET), 5(4), 1830-1833, 2018.
14. Bhavya Ghai. "Analysis & Prediction of American Graduate Admissions Process". Stony Brook University, New York.
15. Dineshkumar B Vaghela, Priyanka Sharma. "Students' Admission Prediction using GRBST with Distributed Data Mining". Gujarat Technological University, Chandkheda.
16. Austin Waters, Risto Miikkulainen. "GRADE: Machine Learning Support for Graduate Admissions". University of Texas, Austin, Texas.
17. https://www.datacamp.com/community/tutorials/k- nearest-neighbor-classification-scikit-learn
18. https://www.meetup.com/Big-Data-Analytics-and- Machine-Learning/events/257926117/
19. Dhruvesh Kalathiya, Rashmi Padalkar, Rushabh Shah., Sachin Bhoite (2019) Engineering College Admission Preferences Based on Student Performance, IJCATR, Volume 8–Issue 09, 379-384, 2019, ISSN:-2319–8656
20. Rohith Gandhi (2018) https://hackernoon.com/introduction-to-machine-learning-algorithms-logistic-regression-cbdd82d81a36