



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 7      Issue: III      Month of publication: March 2019**

**DOI: <http://doi.org/10.22214/ijraset.2019.3470>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**



# Prediction of Individual Student Job Placement on the Basis of Last Year Statistics in Exam and Online Test

Yanni Gupta<sup>1</sup>, Anshuman Tayagi<sup>2</sup>

<sup>1,2</sup>Student Of Electrical And Electronics, Pranveer Singh Institute Of Technology, Kanpur

**Abstract:** Software testing is a major component in modern continuous integration spread environment. Ideally, at every commit, all the system's test cases should be accomplished and also, new test cases should be created for the new code. Supervised machine learning techniques were used to answer the central question: individual student's placement or not, a good predictor of accurately predicting the individual student's placement from the available test. I have a dataset of online test of student's. In this dataset three test results such as quant, reasoning and verbal. In this process dataset prepare firstly by techniques such as classification, clustering, regression, and ranking. And we predict individual student's placement by using various algorithm techniques such as Support vector machine, Random forest, K nearest neighbour, Gaussian Noise, Logistic Regression and Decision Tree.

**Keywords:** Supervised machine learning, random forest, decision tree, Support vector machine, K nearest neighbour, Gaussian Noise, Logistic Regression.

## I. INTRODUCTION

"Ability of a machine to develop its distinguishable routine through the use of a software that employs artificial intelligence techniques to parodist the ways by which humans seem to learn, such as recurrence and involvement". Machine learning has been predictable as essential to the realization of Artificial Intelligence, and it has presentations in various parts of science, engineering and civilization. Machine learning is an solicitation of artificial intelligence (AI) that offers systems the skill to automatically learn and improve from experience without being unambiguously programmed. A core objective of a project is to oversimplify from its capability. Generalization in this perspective is the ability of a learning machine to perform correctly on new, unseen samples/errands after having knowledgeable a learning data set. Process of data analysis comprises requirement analysis, collecting data, clarifying and processing of data, applying algorithms and data results. Classification algorithm is a method that is being used for data exploration. In machine learning and statistics, classification is a supervised learning process in which the computer program learns from the records input given to it and at that point uses this learning to begin new opinion the result shows wide range uncertain factors for the analysis. In this project supervised algorithm such as, random forest, decision tree, Gaussian Noise, Support vector machine, K nearest neighbour, Logistic Regression are moreover used, supervised is a technique which can be used for construction the models that will consign labels to all problem instances. Supervised algorithm are involves minor set of training data for assessing the parameters that are necessary for the classification process. Process of learning by algorithms has controlled to enormous amounts of research in creating algorithms that can be used for prediction. In this project, we work on unindustrialized these algorithms for prediction. Algorithms are planned such a way that give precision. Special consideration is given to detect severe changes in trends. The result shows student's are placed or not.

## II. PERFORMANCE ESTIMATION

### A. Generalization Performance vs. Model Selection:

"First, we feedstuff the training records to our learning algorithm to acquire a model. Second, we predict the makes of our test set. Third, we reckon the number of wrong predictions on the test dataset to compute the model's prediction accurateness. "Depending on our area, though, approximating the performance of a model is not that small, unfortunately. Perhaps we should address the previous question from a different angle: "Why do we care about performance estimates at all?" Ideally, the assessed performance of a model tells how well it performs on unseen data – making predictions on future data is often the main problem we want to solve in applications of machine learning or the expansion of new algorithms. Meanwhile we are usually concerned in selecting the best-



performing model from this set, we necessary to find a way to estimation their respective performances in order to rank them against each other. Going one step beyond mere algorithm perfection, we are usually not only testing with the one single algorithm that we think would be the "finest solution" under the given situations. More regularly than not, we want to relate different algorithms to each other, oftentimes in terms of predictive and computational performance. Let us précis the main points why we assess the predictive performance of a model:

- 1) We want to estimate the generality performance, the predictive performance of our model on coming (unseen) data.
- 2) We want to increase the predictive performance by alteration the learning algorithm and selecting .The best performing model from a given premise space.
- 3) We want to recognize the machine learning algorithm that is best-suited for the problematic at hand;

Thus, we want to equate different algorithms, selecting the best-performing one as well as the best performing model from the algorithm's theory space. Although these three sub-tasks listed above have all in mutual that we want to estimate the performance of a model, they all require different approaches.

### B. Supervised Learning and Classification

This research paper based on supervised learning, a subdivision of machine learning where the target values are known in a given dataset. While many concepts also apply to regression analysis, we will focus on classification, the task of unconditional target labels to the training and test. In classification problems we are trying to predict a distinct number of values. The label commonly comes in unconditional form and represents a finite number of periods.

### C. Learning Algorithm

Another time, our goal is to pearl or projected the target function, and the learning algorithm is a usual of orders that tried to model the target function using a training dataset. A learning algorithm ascends with a hypothesis space, the set of possible premises it can explore to model the unknown board function by expressing the final hypothesis.

### D. Training and testing Method

First, we take a considered dataset and split it into two parts: training and a test set. Then, we fitting a model to the training data and predict the labels of the test set. The division of correct predictions, which can be divided by comparing the predicted labels to the milled truth labels of the test set, constitutes our estimation of the model's prediction precision. Here, it is important to note that we do not want to train and estimate a model on the same training dataset, since it would naturally introduce a very enthusiastic partiality due to over fitting. In other words, we cannot tell whether the model simply learnt the training data, or whether it simplifies well to new, unseen data. (On a side note, we can estimate this professed optimism bias as the modification between the training and test accuracy.) Classically, the terrible of a dataset into training and test sets is a simple procedure of random subsampling. We assume that all data points have been drawn from the same probability distribution (with respect to each class). And we randomly choose 2/3 of these samples for the training set and 1/3 of the samples for the test set.

Stratification

ROLL NO	NAME	QUANT	Reasoning	Verbal	Y/N
1416410001	Akashi Srivastava	12	3	22	1
1416410001	Akashi Srivastava	8	4	17	1
1416410001	Akashi Srivastava	5	5	16	1
1416410001	Akashi Srivastava	6	7	18	1
1416410001	Akashi Srivastava	9	2	18	1
1416410002	Aayush	8	1	20	0



	Mittal				
1416410002	Aayush Mittal	7	2	17	0
1416410002	Aayush Mittal	5	4	14	0
1416410002	Aayush Mittal	8	4	18	0
1416410003	Aayush Singh	11	6	27	1
1416410003	Aayush Singh	10	5	18	1
1416410003	Aayush Singh	5	7	15	1
1416410006	Abhay Yadav	11	4	19	1
1416410006	Abhay Yadav	6	4	18	1
1416410006	Abhay Yadav	11	4	17	1
1416410006	Abhay Yadav	8	6	17	1
1416410006	Abhay Yadav	5	5	21	1
1416410006	Abhay Yadav	7	1	18	1

The designed model will be used further to predict the selection of students who will appear in the OLT's in the upcoming year.

- 1) *Algorithm Use in model:* To the overview routine of our predictive models. We split the dataset into two parts: training and a test dataset. Once the machine learning algorithm fit a model to the training set, we assessed it on the self-determining test set that we withheld from the machine learning algorithm through model fitting. We defined hyper constraints as the parameters of the learning algorithm itself, which we have to require a priori – before model fitting. In contrast, we rose to the parameters of our consequential model as the model parameters. Since the k-nearest neighbours algorithm, one example of a hyper constraint is the integer value of k .If we set  $k = 3$ , the k-nearest neighbours algorithm will expect a class label based on a majority vote among the 3-nearest neighbors in the exercise set. In contrast to k-nearest neighbors, a simple example of a parametric method is logistic regression, a sweeping linear model with an immovable number of model parameters: a weight constant for each feature adjustable in the dataset plus a partiality (or capture) unit. These weight constants in logistic regression, the model parameters, are updated by maximizing a log-likelihood function or minimizing the logistic cost. For appropriate a model to the training. data, a hyperactive constraint of a logistic regression algorithm could be the amount of repetitions or passes over the training set (epochs) in gradient-based optimization. Decision Trees are a type of Supervised Machine Learning the tree can be explicated by two entities, namely decision nodes and leaves. The plants are the decisions or the final outcomes. And the decision nodes are where the data is split. Support vector machines are supervised learning models with accompanying learning algorithms that examine data used for classification and regression analysis.

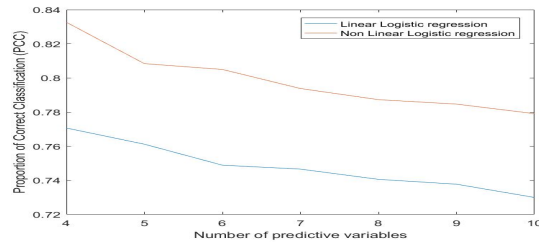


Fig1 logistic regression

Fig2 Gaussian noise

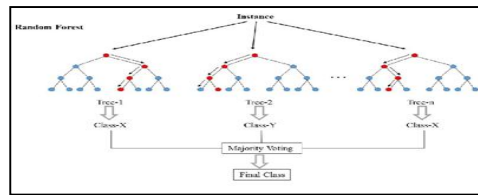


Fig2 Random Forest

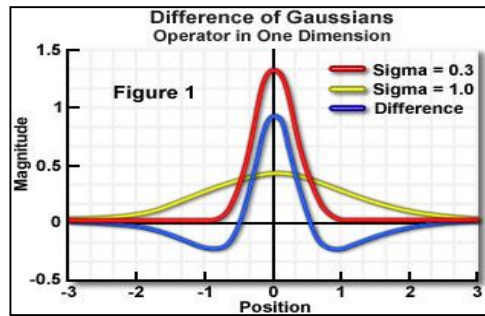


Fig3 Gaussian noise

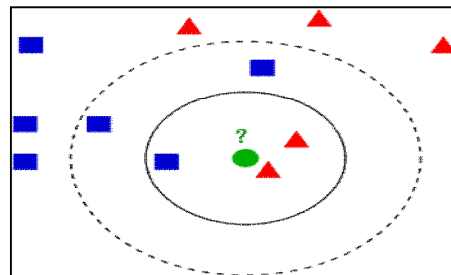


Fig4 k-nearest neighbours

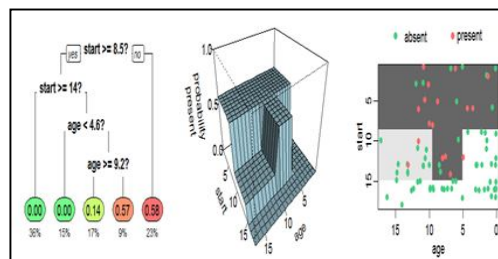


Fig5 Decision tree

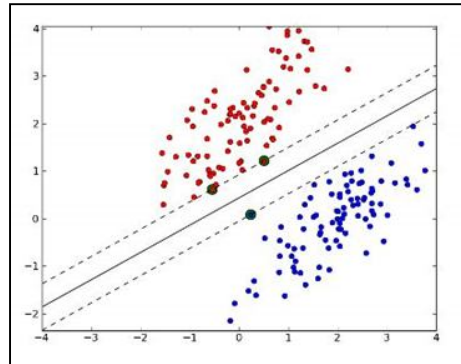


Fig6 support vector machine

2) *Analysed Result:* When model is developed then all algorithm give different accuracy as shown in below:

a) Support Vector Machine (svc): Confusion matrix

[149 274]

[207 2238]

	Precision	recall	f1-score	Support
0	0.88	0.85	0.86	1771
1	0.89	0.92	0.90	2445
age/ total	0.89	0.89	0.89	4216

b) Random Forest Confusion matrix

[1577 194]

[127 2318]

	precision	recall	f1-score	support
0	0.93	0.89	0.91	1771
1	0.92	0.95	0.94	2445
avg/ total	0.92	0.92	0.92	4216

c) K Nearest Neighbour (knn): Confusion matrix

[1536 235]

[144 2301]

	precision	recall	f1-score	support
0	0.91	0.87	0.89	1771
1	0.91	0.94	0.92	2445
avg/ total	0.91	0.91	0.91	4216

d) Gaussian noise: Confusion matrix:

[1509 262]

[243 2202]

	precision	recall	f1-score	support
0	0.86	0.85	0.86	1771
1	0.89	0.90	0.90	2445
avg/ total	0.88	0.88	0.88	4216

e) Logistic Regression: Confusion matrix

[1486 285]

[211 2234]

	Precision	recall	f1-score	Support
0	0.88	0.84	0.86	1771
1	0.89	0.91	0.90	2445
avg/ total	0.88	0.88	0.88	4216

6. Decision Tree: Confusion matrix:

[1620 151]

[158 2287]

	Precision	recall	f1-score	support
0	0.91	0.91	0.91	1771
1	0.94	0.94	0.94	2445
avg/ total	0.93	0.93	0.93	4216

All algorithm gives different accuracy as shown above tables. But Decision tree give 93% accuracy so we take decision tree algorithm result

### III. RESULT

SVM	R_Forest	KNN	N_Bias	L_Reg	D_Tree
1	1	1	1	1	1
1	1	1	1	1	1
0	0	0	0	0	0
1	1	1	1	1	1
1	1	1	1	1	1
0	1	1	0	0	1
0	0	0	0	0	0
1	1	1	1	1	1
0	1	1	0	0	1
0	0	1	0	0	0
0	0	0	0	0	0
0	1	1	0	0	1

Above table shows the results of all algorithms in some cases 2 algorithm fail and 3 pass, vice versa. This result shows 1 and 0 .1 means students selected and 0 means not selected .if 3 algorithm pass any student then that student probability to place because its probability more than 50%.if 2 algorithm passed and in 3 fail then may be student selected or not. This result based on online test result dataset by this predict the student's selection .no algorithm give 100% accuracy but all algorithm give 88-94% accuracy to give right result .

### IV. CONCLUSIONS

As "a picture is substance a thousand words," I want to conclude this paper that summarizes my personal commendations based on the perceptions and literature that was studied. It should be stressed that parametric tests for equating model performances usually violate one or more independent conventions (the models are not independent because the same training set was used, and the estimated overview performances are not independent because the same test set was used.). In an ideal world, we would have admittance to the data predicting the individual student placement on the base of online test result its result to help the improve the performance of students. However, in most practical presentations, the size of the dataset is limited; hence, we can use one of the statistical tests discussed in this article which based on supervised algorithm.



### REFERENCES

- [1] Dweck, C. S., *Mindset: The new psychology of success*, Random House Digital, Inc., 2008.
- [2] WordNet Search, Retrieved from <http://wordnetweb.princeton.edu/perl/webwn?s=mindset>, January 2017
- [3] Shell, M.S., *Thermodynamics and Statistical Mechanics: An Integrated Approach*. Cambridge University Press, 2014.
- [4] Karimi, A., and Manteufel, R. D., "Correlation of Prerequisite Course Grades with Student Performance" ASEE Annual Conference & Exposition, Atlanta, Georgia, ASEE, 2014.
- [5] Taruna, S., and Pandey, M., "An empirical analysis of classification techniques for predicting academic performance", IEEE International Advance Computing Conference (IACC), Gurgaon, 2014, pp. 523-528.
- [6] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software:
- [7] SanavarBangi, PoojaGadakh, PradnyaGaikwad, PratikshaRajpure. "Survey paper on Prediction of Heart Disease Using Data Mining Technique." *International Journal of Recent Trends in Engineering & Research (IJRTER)*, Volume 02, Issue 03; March - 2016 [ISSN: 2455- 1457].
- [8] AnchalGoyal, RanpreetKaur. "Loan Prediction Using Ensemble Technique." *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 5, Issue 3, March 2016.
- [9] R. Camilleri, D.A. Howey, M.D. McCulloch. "Predicting the Temperature and Flow Distribution in a DirectOil-Cooled Electrical Machine with Segmented Stator." *IEEE Transactions on Industrial Electronics* · January 2015
- [10] Mohamed Krid, FaizBenamar, and Roland Lenain. "A new explicit dynamic path tracking controller using Generalized Predictive Control." *International Journal of Control, Automation and Systems*.





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)