Ensemble method to predict impact of student intelligent quotient and academic achievement on placement

Kanika Thakur
Research Scholar
Birla Institute of Technology,
Patna, Bihar, India
kanikaswap@gmail.com

Dr Kanhaiya Lal
Assistant Professor & In-charge
Computer Science
Birla Institute of Technology,
Patna, Bihar, India
klal@bitmesra.ac.in

Vinay Kumar
Assistant Professor, CSE
Amity University,
Greater Noida, India
vkumar@gn.amity.edu

2021 2nd International Conference on Intelligent Engineering and Management (ICIEM) | 978-1-6654-1450-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICIEM51511.2021.9445323

***Abstract*: - The study's aim is to see how academic achievement and student Intelligence Quotient influence placement. This paper will attempt to predict whether a student's intelligence quotient or academic score plays a significant role in placement. On a dataset of 193 students, we used a machine learning algorithm to compare the impact of student intelligence, behavior, and academic achievement on placement. We have used a Voting Classifier architecture to predict and classify the probability of a student being placed or not. The motivating force behind this research was to figure out why a group of students scoring the same marks in the same branch studying under the supervision of the same faculty are not able to fulfill the demands of an organization in order to be employed. The aim of this research was to combine conceptually different machine learning classifiers and predict the probability of a student being hired using a majority vote or the average expected probabilities. A classifier like this can be useful for balancing out the weaknesses of a group of models that are all performing well. Experiments show that student intelligence and attitude play a significant role in the recruiting process.**

Keywords:-Machine Learning, Intelligent Quotient, education data mining, VotingClassifier.

I. INTRODUCTION

In recent years, there has been tremendous work done in different fields of education data mining were different algorithm are applied to find the hidden information from the student database and the overall objective of the research was to find the factors that can be further utilized by the institute to ensure the overall success of students and help them to overcome the challenges of their life. A significant number of studies have explored the associations between academic accomplishment and various psychological structures such as self-concept [1, 2], attitude [3], and emotional intelligence. Studies have demonstrated that children with varied backgrounds in the school context respond accordingly [4, 5, 6] Spearman discovered that children tended to achieve similar scores in unrelated school subjects, and he hypothesized that their outcomes were influenced by a common factor. The use of factor theory [7] revealed that there is a single fundamental general ability factor that accounts for performance on various types of tests. Others [8, 9] disagree, arguing that this aspect has no effect on results on assessments of basic cognitive abilities. What exactly is this? “Intelligence is described as the ability to manipulate information (tasks that require more cognitive complexity).” While no single test can measure general brain ability, the findings of research can be used to assess more specific cognitive skills.

II. HISTORICAL CONTEXT AND RELATED WORK

The hypothesis that General Mental Ability (GMA) is a good predictor of job success is supported by thousands of test results. Based on the nature of the job and how output is measured, GMA explains between 30% and 70% of the difference in individual work success (i.e. correlations of.56 to.84), which is higher than any other known predictor. Before delving deeper into this evidence, we need to understand what we mean by improvement. There are three main indicators of job success. In the hierarchy of occupations, performance evaluation on tasks similar to those encountered on the job (work-sample tests), and quality evaluations from managers are placed. The outcomes of supervisors' work-sample behaviors and grades are forecasted by GMA. When output is evaluated using work-sample tests, the correlation between GMA and progress is 0.84, according to data from several meta-studies. When supervisor ratings are used, the relationship is smaller, at 0.74 for high-complexity workers[4]. GMA also predicts how much you advance in your career, or how much work you do (Dawis, 1994; Jensen, 1980, pp. 339– 347).” Schmidt, Frank L., and John Hunter Schmidt, Frank L., and John Hunter Hunter, John Hunter Hunter, John Hunter Hunter, John Journal of personality and social psychology 86.1 (2004): 162. "General mental capacity in the world of work: occupational attainment and job performance." According to data from the US Jobs Service, there is a strong connection (0.72) between GMA and occupational level, and US military data shows that GMA ratings are higher at higher levels of employment. At lower stages of a work, there is also a wider variety of GMA scores than at higher stages. There tend to be high-scoring individuals in low-level jobs, but low-scoring individuals cannot be promoted to higher ranks[4].

Experts are still debating the definition of intelligence, but it does appear to be a primary indicator of learning success in some studies.

[10,11,12] Intelligence is thought to be a significant variable that separates the extremely gifted from the extremely poor, as well as a variable that distinguishes standards of academic achievement [13, 9] and [14]. However, as schooling progresses, there is a steady decline in correlation coefficients, implying the value of other variables such as socio-familial, informal, and school-related[15]. Teachers also use school grades from school reports to select students for different secondary school tracks[16]. According to recent studies, teachers can still use positive or negative student progress data when determining placement guidelines. Caro et al. (2009) discovered that students who were growing[17] faster in their mathematical abilities as measured by standardized performance tests were more likely to earn a high-track recommendation than students who were making slower progress. In a study of pre-service teachers, K. Lapproth and Fischer discovered that students who strengthened their school marks[18] in the last year of primary school were more than twice as likely to be recommended for the highest grade in secondary school than those whose school marks declined, despite having the same overall grade. The effect of student achievement growth on teachers' school-placement decisions can be explained by suggesting that teachers are motivated by expectations that they have based on previous student development. Cooper and coworkers [19] coined the term "sustaining expectation effects" to describe the tendency for teachers to expect students to continue to succeed based on previously established patterns of achievement. Students who performed well in the past are expected to perform well in the future, while students who performed poorly in the past are expected to perform poorly in the future, according to the theory of sustained expectation effects.

III. Description Of Working Model

It has been a source of long debate and study to determine the factors that influence a student's career advancement and work placement. One of the factors is academic score, which has been extensively studied, and the other is intelligence score, which has received less attention.

As a result, the void is filled by determining the effect of academic performance on placement as well as the relationship between Intelligent Quotient and student placement. We used a dataset of 193 instances, which represents the number of BTECH, MBA, MCA, BCA, and B.SC(IT) students at Amity University. The Student Id, Student Name, 10th, 12th, Graduation, and Placement status are all included in this dataset. The sample dataset for evaluating the influence of academic achievement and intelligent quotient is made up of students from various universities in Jharkhand who volunteered for this report. Since it was not right to enter the students' incorrect scores, we eliminated the noise by binning method. After that, we had the dataset shown in figure 1.The first and most important step was to determine the student's intelligent quotient. For this, the WIAS-IV-IQ test was used. As this method is often used to assess intelligence and cognitive ability in adults and older adolescents. For assessing the IQ , software was developed, that allowed to register students and administer the test. The test consisted of 10 core subtests and five supplemental subtests, with the 10 core subtests measuring intelligence, and cognitive ability. The Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), Processing Speed Index (PSI), and Full Scale IQ (FSIQ) are four index scores that represent major components of intelligence. They were calculated using the cumulative combined output of the VCI, PRI, WMI, and PSI.

The IQ test module comprised of five tests, each with a time limit. The score obtained after the test was conducted on students are as shown in Tables 1 and 2, each question set consisted of 30 questions.

- Test Score Obtained

STUDENT_	ST_MARKS	TEST1	TEST2	TEST3	TEST4	TEST5
ST11	80	87.86	88.86	87.86	105	10
ST12	87.86	89.97	71.4	71.4	92.97	94.5
ST13	71.4	87.41	90.03	90.03	87.41	89.4
ST14	93.67	91.05	105.84	105.84	91.05	93.0
ST15	78.83	94.69	86.87	86.87	94.69	96.6
ST16	60.27	76.49	118.49	118.49	76.49	78.4
ST17	67.69	87.41	102.68	102.68	87.41	89.4
ST18	97.39	112.88	105.84	105.84	112.88	114.8
ST19	75.12	138.36	109.01	109.01	138.36	140.3
ST20	67.69	83.77	96.36	96.36	83.77	85.7
ST21	104.81	123.8	99.52	99.52	123.8	125
ST22	89.96	87.41	105.84	105.84	87.41	89.4
ST23	71.4	94.69	118.49	118.49	94.69	96.6
ST24	71.4	91.05	83.71	83.71	91.05	93.0
ST25	75.12	94.69	102.68	102.68	94.69	96.6
ST26	67.69	87.41	86.87	86.87	87.41	89.4
ST27	108.52	127.44	83.71	83.71	127.44	129.4
ST28	115.95	134.72	127.98	127.98	134.72	136.7
ST29	67.69	87.41	121.66	121.66	87.41	89.4
ST30	71.4	91.05	93.19	93.19	91.05	93.0
ST31	75.12	94.69	83.71	83.71	94.69	96.6
ST32	82.54	94.69	99.52	99.52	94.69	96.6
ST33	67.69	87.41	102.68	102.68	87.41	89.4
ST34	78.83	98.33	102.68	102.68	98.33	100.3
ST35	67.69	87.41	105.84	105.84	87.41	89.4
ST36	75.12	91.05	105.84	105.84	91.05	93.0
ST37	97.39	94.69	109.01	109.01	94.69	96.6
ST38	78.83	98.33	83.71	83.71	98.33	100.3
ST39	86.25	101.97	105.84	105.84	101.97	103.5
ST40	63.98	105.6	71.06	71.06	105.6	107
ST41	67.69	87.41	105.84	105.84	87.41	89.4
ST42	82.54	101.97	64.73	64.73	101.97	103.5
ST43	78.83	101.97	96.36	96.36	101.97	103.5
ST44	78.83	127.44	102.68	102.68	127.44	129.4
ST45	67.69	127.44	134.31	134.31	127.44	129.4
ST46	86.25	101.97	109.01	109.01	101.97	103.5
ST47	86.25	105.6	109.01	109.01	105.6	107
ST48	71.4	101.97	105.84	105.84	101.97	103.5

Table 1

Total no of Student

```
<class 'pandas.core.frame.DataFrame'
RangeIndex: 193 entries, 0 to 192
Data columns (total 7 columns):
STUDENT_ID     193 non-null object
ST_MARKS       193 non-null float64
 TEST1         193 non-null float64
TEST2          193 non-null float64
TEST3          193 non-null float64
TEST4          193 non-null float64
TEST5          193 non-null float64
dtypes: float64(6), object(1)
memory usage: 10.6+ KB
```

Table 2

The mean score of all the test, standard deviation and min IQ score that ranges from 60 to 77, Maximum IQ ranges from 120 to 139 and the most significant part is that 75% of the student has scored between 90 to 106 as expected but we see there is a significant difference between the score obtained. In figure 1, 2, 3,4,5,6 shows the distribution of IQ Score obtained for various students in context of different test given by an individual student.

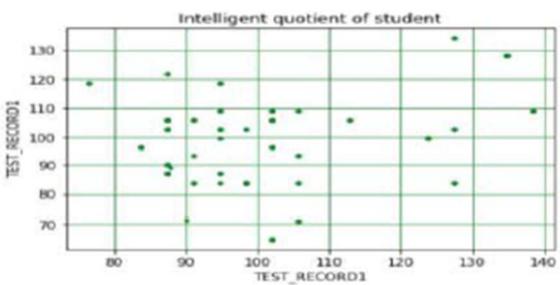

Figure 1

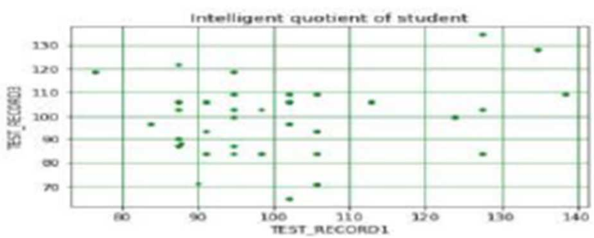

Figure 2

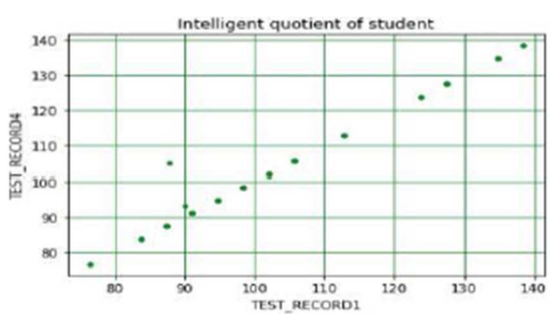

Figure 3

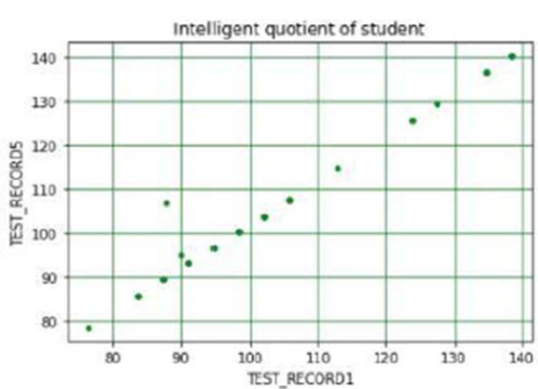

Figure 4

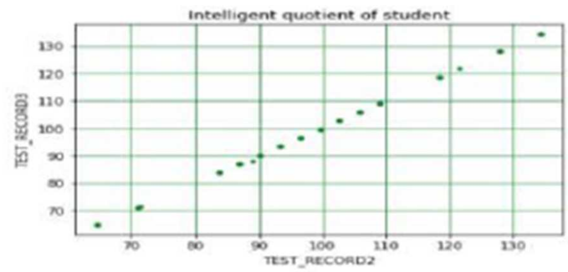

Figure 5

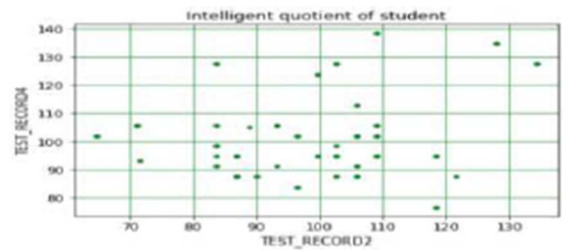

Figure 6

After obtaining the IQ score from the five tests twice for an individual student, The **DBSCAN** clustering method is used, Here, IQ = {IQ1, IQ2, IQ3, …,IQn} are the set of data points. Here, data points are the IQ score of the students and act as core point if the circle around it contains at least minimum points 4, the epsilon value we have taken is 7, the cluster points are represented in red color. Those IQ that does not fall as borderpoint, and if there are no other IQ score obtained by the students that doesn't falls within epsilon radius, then it treated as Noise. As shown in the figure 7. It groups highly similar IQ score of a student into a single cluster, using the concept of epsilon radius around every data points and classifies them into Core point, border point and a noise. In this way we remove the outliers from the dataset, the IQ score obtained by individual students which are the part of the cluster becomes the IQ score that is accepted for further being considered as to calculate the standard IQ score of the student as shown in eq1.

Standard Score is calculated as :-

$$z = (x - \mu) / \sigma \qquad (1)$$

Where, μ is the population mean and σ is the population standard deviation.

If more, the two Clusters are formed in that case, the silhouette score of the clusters are calculated, and the one with higher silhouette score is considered as the best cluster and the IQ score that falls within the best cluster are fetched, and used further to calculate the standard IQ score of the students using the formula discussed in eq1. The formula used to calculate the distance, Silhouette score and IQ score are mentioned in equation 2 & 1

$$\text{Silhouette score} = (p - q) / \max(p, q) \qquad (2)$$

Here, p = mean distance to the points in the nearest cluster, q = mean intra-cluster distance to all the points.

The clusters obtained after applying the DBSCAN algorithm, on the IQ score of student, with student Id st11, are shown in figure 7. Here the IQ score in the cluster are marked with red color.

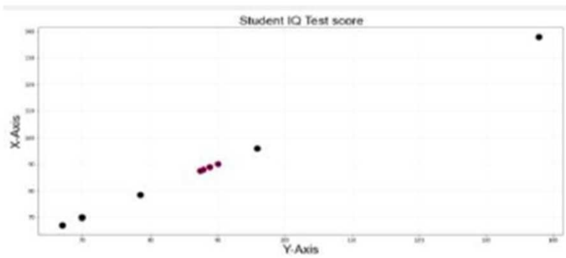

Figure 7

After the standard IQ has been obtained, ensemble methods is used to assess the influence of academic achievement and student IQ on student success and Placement. The entire dataset was partitioned into 80:20 ratio. The 80 percent of the dataset is used to train the model and 20 percent of the dataset is used to test the model. We have used 2-layered ensemble model to assess the influence of academic achievement and student IQ on Placement. The first algorithm is referred to as the Base learner, while the second is referred to as the Meta learner. The Meta-learner processes all of the Base-prediction impacts.

We've used a diverse and dependable heterogeneous system known as stacking. Each layer consists of one or more templates, and each subsequent layer gets more knowledgeable about the prior one. For solving the binary classification problem, we have used random forest classifier. Random forest classifier are based on decision trees, and are both excellent ensemble algorithms, so they were employed in combination to classify the sample instances. RF does not need to make assumptions about data distribution and can process thousands of variables without loss of it. However, the structure of RF is not robust and tends to overfit the training data, and thus it's generalization suffers. To fix the overfitting problem, a decision tree classifier is used as the third model in the training set. Bagging works particularly well when the learner's data is highly variable and has significant predictive power; for instance, minor alterations in the input results in large changes in the output. It decreases the variance by aggregating the various statistical properties of the learners into a more homogeneous population. High-variance models such as Decision Trees perform admirably. We use voting classifier in our sample as the Meta learner.

IV. Result Analysis

In this part, we havediscussed the results we have obtained by doing analysis on our student's dataset. The accuracy achieved by using the Random Forest Classifier for student placement prediction was 92 percent, and the accuracy achieved by using the Bagging Classifier for student placement prediction was 95 percent, and the meta learner for which voting Classifier for student placement prediction was 94 percent. To test the efficacy and stability of the proposed ensemble model, ten prediction experiments were conducted. As shown in Table 3, the prediction results were obtained after performing the 10 rounds of the proposed ensemble model. The average value of the 10 rounds was used to calculate the overall prediction efficiency of the proposed ensemble model. 0.9700, 0.9500, 0.96, 0.1460, and 0.9185 arsion, recall, and F1 vaues.

Test results of the proposed ensemble model from 10 runs.

Run Index	Class Label	Precision	Recall	F1
1	0	0.9400	0.9900	0.9600
	1	0.9700	0.9400	0.9500
2	0	0.9400	0.9700	0.9600
	1	0.9600	0.9400	0.9500
3	0	0.9500	0.9700	0.9600
	1	0.9600	0.9600	0.9500
4	0	0.9400	0.9600	0.9500
	1	0.9500	0.9400	0.9400
5	0	0.9400	0.9700	0.9600
	1	0.9600	0.9400	0.9500
6	0	0.9400	0.9900	0.9600
	1	0.9700	0.9400	0.9500
7	0	0.9500	0.9700	0.9600
	1	0.9600	0.9600	0.9500
8	0	0.9500	0.9900	0.9700
	1	0.9900	0.9400	0.9600
9	0	0.9500	0.9900	0.9600
	1	0.9700	0.9500	0.9600
10	0	0.9500	0.9900	0.9700
	1	0.9700	0.9500	0.9600
Min		0.9400	0.9400	0.9400
Max		0.9900	0.9900	0.9700
Ave		0.9550	0.9600	0.9565

Table 3

This segment assesses the significance of each function in predicting a student's placement. The feature extraction method choose seven, features from a total of ten, which were then chosen for further testing. Three classifiers (RF, Bagging, Decision Tree, and Voting Classifier) are used in experiments to examine the impact of each function on predicting student success. Using the Decision Tree classification process, we discover that the "Intelligent quotient" is the best predictor of the desired student's success.

By comparing the performance of the best and worst features, the proposed proposition focused on the Intelligent Quotient and Placement Performance features appears to be useful in enhancing classification accuracy. When we compare the impacts of other features to intelligent quotient features, we find that intelligent quotient has greater accuracy. As a result, it can be inferred that students with higher IQ tend to perform better in the placement process and are placed well.

V. CONCLUSION

This study was set out to explainhow data can help fight dropout. Many algorithms have come into play and worked with a significant detail from a simple dataset to a dataset of greater complexity. Compared to a meta-classifier such as Random Forest, classifier, and bagging, and the Ensemble Stacking Classification approach performed better, resulting in better Placement predictions. By integrating those three algorithms, this approach can achieve a precision of 94% and a recall rate of 95%. The Ensemble Stacking Method's ability to predict student placement is sufficient. We also discovered that different characterizations of features influence prediction, including the percentage of students that attends, assignment grades, GPA, parental income, parental age, and gender affect student performance. However, it has been shown that students are positioned differently according to academic success, study plan, and climate variables. For future work, we recommend increasing the number of correlative features as well as well as the size of the dataset, which would help with performance, so that will be an external evaluation.

REFERENCES

[1] Susperreguy, Maria Ines, Pamela E. Davis-Kean, Kathryn Duckworth, and Meichu Chen. 2017. "Self-Concept Predicts Academic Achievement Across Levels Of The Achievement Distribution: Domain Specificity For Math And Reading". *Child Development*

[2] Wolff, Antonio C., M. Elizabeth Hale Hammond, Kimberly H. Allison, Brittany E. Harvey, Pamela B. Mangu, John M.S. Bartlett, and Michael Bilous et al. 2018. "Human Epidermal Growth Factor Receptor 2 Testing In Breast Cancer: American Society Of Clinical Oncology/College Of American Pathologists Clinical Practice Guideline Focused Update". Journal Of Clinical Oncology 36 (20): 2105-2122. doi:10.1200/jco.2018.77.8738.

[3] "Janošević, M., &Petrović, B. (2018). Effects of personality traits and social status on academic achievement: Gender differences. Psychology In The Schools. doi: 10.1002/pits.22215"

[4] Schmidt, Frank L, and John E Hunter. "Select on intelligence." Handbook of principles of organizational behavior(2000): 3-14.

[5] Schmidt, Shaffer, and Oh, 2008). From: Schmidt, Frank L, and John E Hunter. "Select on intelligence." Handbook of principles of organizational behavior(2000): 3-14.

[6] Min M. C., Islam M. N., Wang L., Takai J. (2018). Cross-cultural comparison of university students' emotional competence in Asia. *Curr. Psychol.* 10.1007/s12144-018-9918-3

[7] Gottfredson, Linda S. The general intelligence factor.Scientific American, Incorporated, 1998.↩

[8] Horn, John, and Hiromi Masunaga, Charness, Neil, Paul J Feltovich, and Robert R Hoffman, 2006

[9] L.S. Almeida, M.A. Guisade, R. Primi, G. Lemos. Contribuciones del factor general y de los factoresespecíficos en la relación entre inteligencia y rendimiento escolar. European Journal of Education and Psychology, 1 (2008), pp. 16-25"

[10] Spinath, Birgit, Frank M. Spinath, Nicole Harlaar, and Robert Plomin. 2006. "Predicting School Achievement From General Cognitive Ability, Self-Perceived Ability, And Instrinsic Value" Intelligence 34 (4): 363-374 doi:10.1016/j.intell.2005.11.004.

[11] Sternberg, Robert J.; Grigorenko, Elena L.; and Bundy, Donald A. (2001) "The Predictive Value of IQ," *Merrill-Palmer Quarterly*: Vol. 47 :Iss. 1 , Article 2. Availableat: https://digitalcommons.wayne.edu/mpq/vol47/iss1/2

[12] Deary, Ian J., Steve Strand, Pauline Smith, and Cres Fernandes.2007."IntelligenceAndEducational Achievement". Intelligence 35 (1): 13-21 doi:10.1016/j.intell.2006.02.001.