The role of predictive analytics to explain the employability of management graduates

Ramakrishnan Raman
Symbiosis Institute of Business Management Pune,
Symbiosis International University, Pune, India, and
Dhanya Pramod
Symbiosis Center for Information Technology, Symbiosis International University,
Pune, India

Abstract

Purpose – In India, one of the prime focuses of a post-graduate management program is to prepare students and make them job-ready. Masters in Business Management (MBA) program helps students to imbibe theoretical and practical skills which are required by the industry, which can make them hit the ground running from the day they start their career. Many students (almost 40–50%) get pre-placement offers based on their performance in summer internship. The selection for summer interns by the corporate happens within a few months of the student joining the MBA program. Signaling theory in education indicates that the level of productivity of an individual is independent of education, but the educational qualification acts as a testimony for higher ability. However, this theory does not explain the reason for the mismatch between “education and work” or “education and the disparity in salary” between individuals who earn differently but have the same qualification. The paper aims to explore three attributes namely – “employability” – the chance of being employable; “pre-placement offers” – the chance of securing a job offer based on the performance in internship and “salary” – the chance of bagging a good job offer with a high salary.

Design/methodology/approach – The authors have used longitudinal data consisting of 1,202 students who graduated from reputable business schools (B-Schools) in India. In the study, the authors have used predictive analytics on six years data set that have been gathered. The authors have considered 24 attributes including educational background at the graduate level (B.E, B Tech, B Com, BSc, BBA and others), score secured in class ten (high, medium and low), score secured in class twelve (high, medium and low), score secured in graduation (high, medium and low), competency in soft skills (high, medium and low), participation in co-curricular activities (high, medium and low) and social engagement status (high, medium and low).

Findings – The findings of the study contradict the signaling theory in education. The findings suggest that the educational qualification alone cannot be the predictor of the employability and the salary offered to the student. The authors note that the better performance at a lower level of qualification (class 12) is the strong predictor in comparison to the student performance at their graduation and post-graduation level. The authors further observed at the post-graduate management education level that soft skills and participation in co-curricular activities are the major deciding factors to predict employability and pre-placement job opportunity and marks secured in class 12 is one more factor that gets added to this list to predict salary. The paper can immensely help management graduates to focus on key aspects that can help to hone appropriate skills and also can help management institutions to select the right students for management programs.

Research limitations/implications – The analysis and the predictive model may apply to Indian B-Schools wherein the quality of students are almost the same or better. Predictive analytics has been used to explain the employability of management graduates alone and not any other.

Practical implications – The authors’ study might be useful for those students who often fail to understand “what” skills are the most important predictors of their performance in the pre-placement and final-placement interviews. Moreover, the study may serve as a useful guide to those organizations that often face dilemmas to understand “how” to select an ideal candidate for the particular job profile from a campus.

Originality/value – The authors believe that the current study is one of the few studies that have attempted to examine the employability of management graduates using predictive analytics. The study further contradicts that the signaling theory in education does not help better explain the employability of the students in extremely high-paced business environments.

Keywords Employability, Predictive modeling, Predictive analytics, Management education, Machine learning

Paper type Research paper
1. Introduction
The B-Schools in India can be classified as Tier -1, Tier -2 and Tier -3. The Tier -1 is of the top B-schools of the country, which have an excellent infrastructure and also offer good quality education to students and there is a bee-line queue of students who aspire to join them. Corporates through these B-Schools and give excellent job opportunities to students pursuing their management program in them. A very limited number of students to get an opportunity to pursue their management program in these B-Schools. They do not have enough seats for all meritorious students. The Tier-2 and Tier-3 schools are those which manage to get students to join them by showcasing that they are as good as Tier -1. The number of applications that these B-Schools attract is significantly lower than the Tier -1 B-Schools.

There are several competitive examinations in India, which is the first step for those who aspire to pursue an MBA. The popular ones are common aptitude test (CAT), Xavier aptitude test (XAT) and Symbiosis national aptitude test (SNAP) and a few others. CAT is a computer-based exam and the test scores a candidate on the bases of quantitative ability, verbal ability and reading comprehension, data interpretation and logical reasoning. CAT is the entrance exam for admission to 20 Indian Institutes of Management (IIMs) located all over India. XAT is an all-India test conducted by Xavier Association of Management Institutes (XAMI). This exam is the gateway for admission to the post-graduate programs in management Xavier Labor Relations Institute, Xavier Institute of Management and a few other renowned private B-Schools. SNAP is a national-level management entrance exam, which is administered by Symbiosis International (Deemed University) (SIU) – for admission in Symbiosis Institute of Business Management and other Symbiosis institutes which offer MBA programs. These exams use both quantitative and qualitative criteria for selecting students. Apart from the score, performance (grades/marks) in the undergraduate program and proficiency in communication skills and other qualitative aspects are evaluated through written statements of purpose, group discussions/group exercises and personal interview process.

After admission, students are put through a rigorous coursework that helps in enhancing their knowledge and skills. Every organization which hires students prefer only the best talent to work for them. Hence, B-schools are always cautious about their admissions strategy, which would help them choose the right candidate for their program. In India, if a B-School fails to provide its students with appropriate job opportunities, it generally has a cascading effect on its ability to attract the crème de la crème students.

A post-graduate program in management or business administration helps in holistic development of students by equipping them with the knowledge and skills required by the corporate world. Such programs provide students an exposure to real world case studies, simulation and industry training along with the academic theories which can be applied in industry. It also helps them in making them industry leading professionals.

This makes MBA to be a sought-after course for aspirants who aspire to start their career in the best corporations. In India, many B-Schools offer a two-year MBA program. The program is structured in such a manner that the students take their summer internship at the end of their first year and then take the final job offers by the end of their second year.

There have been several empirical studies that have been conducted to identify changes to be made in MBA education that can help in hiring MBA graduates. The literature gives very little insights into the attributes and its correlation to employability. Inspite of the rigorous selection process, a few students get lucrative job offers and a few struggle for the same. Literature does not give any insights into this aspect. Although signaling theory in education indicates that the educational qualification acts as a testimony for higher ability, the aspects that impact employability of management graduates has not been covered in the existing literature. This was the prime motivation for this study.

The gap identified led to the research questions – including which are the key aspects and attributes influence the chance of getting a job immediately after pursuing an MBA? Which
are the key aspects that influence the chance of bagging a high salary (pay package) after pursuing an MBA? Which are the key aspects that influence the chance of bagging a pre-placement job offer (job offered based on performance in internship) after pursuing an MBA? Answering these questions can immensely help management graduates to focus on those key aspects which can help them hone the appropriate skills and also can help management institutions to select appropriate students for their management programs and this will be a significant contribution of this study. Predictive analytics has been used to find the key attributes that have an impact on employability of management graduates. The research questions have been answered using data and predictive analytics.

2. Literature review

2.1 Management education in India

It was in 1953 when the first B-School named the Indian Institute of Social Welfare and Business Management was set up in Kolkata. This B-School was set up jointly by the West Bengal Government and Kolkata University. It was in 1955 when Delhi University started offering an MBA program. In the same year, two more universities namely, Madras University and Bombay University started offering MBA. Later in 1957, Andhra University also started to offer an MBA program. In 1960, to promote and build excellence in management education the Government of India established IIMs. The first IIM was set up in Kolkata in 1961 and the second one was set up in 1962 at Ahmedabad. Many more IIMs were established and in 2021 there are 20 IIMs in India, which are offering management education. Over the years, there have also been many private B-Schools that were established and started competing with the IIMs to offer quality management education. Figure 1 shows the key milestones of MBA education in India.

Premier Indian B-Schools attract high-caliber students and also are the sought-after place for corporate. Literature indicates that education provided in B-Schools must reduce the gap between the industry and academia (Sanjeev and Dash, 2011). While the real outcome of management education is not easy to determine (Longenecker, 1997), one prime aspect of management education is to create leaders for the corporate world. In order to remain competitive, any organization has to focus on imparting management education to its work force or hire management graduates who can help them in doing the needful (Longenecker and Ariiss, 2002). In the Indian context, management education is also about imparting the aspects related to job-readiness, which can help the individual to contribute to the growth and long-term performance of the corporate (Krishnan, 2008). The literature indicates that many Indian B-Schools focus on increasing the salary packages of graduates. While analytics has been used by a few Indian B-Schools to predict the faculty attrition, using it for predicting placement is not indicated in the literature (Raman et al., 2019). There is significant literature that indicate the need to restructure and improve management education (Bennis and O’Toole, 2005; Donaldson, 2002; Emiliani, 2006; Etzioni, 2002; Ferraro et al., 2005; Ghoshal, 2003; Hamel, 2009; Mintzberg et al., 2002; Pfeffer and Fong, 2002; Tsurumi, 2005). Many of them indicate the need to impart relevant knowledge and skills needed to make the graduate employable and also help them in critical thinking. Literature indicates that attaining excellence in imparting management education is a challenge (Acito et al., 2008; Cornuel, 2005; Friga et al., 2003; Hamel, 2009; Hawawini, 2005), but ensuring that students are given the right set of knowledge and skills along with the focus of making them employable are unsaid mandate for any management program. In India, there are several thousand B-Schools and they are classified as “good” and the “not so good.” The Government of India has launched the National Institutional Ranking Framework (NIRF), which is a methodology to rank Indian B-Schools. Since 2015, this framework has been adopted by the Indian Ministry of Education. This framework helps in classifying the B-Schools as the good and the not so good based on
several parameters. While there are prestigious B-Schools that offer the best quality education, there are many B-Schools that do not even fill their seats. The corporate has their concern regarding business education imparted by these B-Schools especially due to the quality – as many of them do not serve the needs of the corporate world (Elmuti, 2004; Blass and Weight, 2005; Thomas, 2007).

A B-School offering management education can achieve excellence when it helps in creating value (both financial and non-financial) to the student which can, in turn, help in balancing social and economic interest of the individual and also the society at large. Excellence in business education is about creating value (not only financial), contributing to sound economic growth, improving human conditions and balancing social and economic interest (Emiliani, 2004). In India, one major aspect that remains a strong influence for an aspirant in selection of a B-School for pursuing an MBA is the “Job opportunities” that become available to him/her.

Machine learning has the ability to revolutionize management education by increasing efficiency, using learning analytics, implementing predicative analytics, adaptive learning, personalized learning and in-assessment. This has been tabulated as shown in Table 1.

2.2 Employability of management graduates
A management program must impart values and offer a deep understanding of management concepts that are a blend of theory and practice, which can help students to transform themselves as better individuals who can transform corporate and help them in achieving
their goals (Burns, 1978; Krishnan, 2008). An individual’s ability to get a job that he/she desires is “employability” (Rothwell and Arnold, 2007). Employability is an individual construct (Forrier et al., 2018) and is a vital instrument for the modern labor market (Forrier and Sels, 2003). Another definition of employability is the individual’s perception about his/her ability to find and continue employment (Vanhercke et al., 2014). While employability is specific to an individual that leads to a positive future outcome (Forrier et al., 2018), literature indicates (McArdle et al., 2007) that management education has to improve the individual’s perception of employability and also the actual potential of employability. While there are several aspects which influence employability, key aspects across domains of human capital, psychology, sociology and behavioral science are critical for achieving success in a career (Hirschi et al., 2018).

It is a globally accepted phenomenon that management education has to help in imparting skills and knowledge to enhance the “job-readiness” of graduates and make them employable (Moore and Morton, 2017). In the Indian context, students pursue an MBA program that will help them to find a job or would help them in improving their existing employment status.

Indian corporate searches for B-School graduates who can be employed, who will make valuable contributions to their organization. There is enough research literature which shows that there is a gap between the skills, knowledge and abilities of business graduates and the needs of Indian businesses. Research literature also indicates that apart from providing “domain knowledge” Indian B-Schools must also focus on aspects that will help management graduates perform well in the job that they take up after graduation (Baruch et al., 2019). During the 1960s and 1970s, the graduates from B-Schools met the immediate requirements and needs of the corporate that had very specific and defined functional specialties (Miles, 1985). In the present day context, this does not happen. A research conducted by the Associated Chambers of Commerce of India (ASSOCHAM), which is one of the oldest and largest apex chambers of commerce and industry of India, revealed that only 7% of MBA graduates are employable (Assocham, 2016).

Existing research indicates that there are several skills that an MBA graduate must have in order to be “employable” by industry. Some critical skills required for employability include excellent communications skills, very good interpersonal skills, ability to solve complex problems, ability to work in a team, having good emotional intelligence, being creative in finding solutions to critical business problems and good work ethics (Cherniss et al., 1998; Crosling and Ward, 2002; Friga et al., 2003; Jackson, 2010; Pinard and Allio, 2005;

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Application</th>
<th>Employability of management graduates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning analytics</td>
<td>Learning analytics can help teachers to gain insight into data that cannot be gleaned by using the human brain. This can positively impact the teaching and learning process</td>
<td></td>
</tr>
<tr>
<td>Predictive analytics</td>
<td>Predictive analytics can help in arriving conclusions about future events. Example: Using a data set of graduates students’ cumulative records, predictive analytics can help in informing about those students who are more likely to drop out because of academic failure</td>
<td></td>
</tr>
<tr>
<td>Adaptive learning</td>
<td>Adaptive learning can be used to remediate struggling students or challenge gifted ones. Adaptive learning is a technology can help analyze a student’s performance in real time and can help in modifying teaching methods</td>
<td></td>
</tr>
<tr>
<td>Personalized learning</td>
<td>Personalized learning can help in giving each student an individualized learning experience. Personalized learning can help in enabling a student to go at their own pace and, in some cases, making their own decisions about what to learn</td>
<td></td>
</tr>
<tr>
<td>Assessment</td>
<td>Machine learning can be used to grade student assignments and exams more accurately than a professor. It may require some input from the professor, but the results will have higher validity and reliability</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Machine learning applications in management education
Bharambe et al., 2017). Research literature also indicates that enhancing the employability skills of MBA graduates can immensely help in bridging the between industry expectations and academic reality (Abrahamson et al., 2016; Agarwal et al., 2019; Hughes et al., 2011; Taylor and Hooley, 2014). Demographics and employment history also has a role to play in employability.

Although several researchers have suggested the essential skill set required for an MBA graduate to become employable (Ayoubi et al., 2017; Benson et al., 2014) and many studies have examined the gaps and lacuna that exists in the Indian MBA program (Pandiyan, 2011; Bansal and Agarwal, 2019; Bhatia and Panneer, 2019), there is no concrete analytical and quantitative evidence that explores and identifies specific factors that affect employability of a management graduate. A machine learning (ML) model can immensely help in identifying the factors that influence employability (Bharambe et al., 2017; Piad et al., 2016; Othman et al., 2018; Alghamlas and Alabduljabbar, 2019).

2.3 Machine learning models for prediction
ML is an application of artificial intelligence (AI) wherein the systems learn from data using algorithms. These insights are used for predictions on the unseen dataset that are similar. Machine learning algorithms are of two types, namely, supervised and unsupervised. Many ML techniques have evolved based on mathematical or statistical models. Classification and regression are two of the most common analyses for which ML models are used. Classification denotes a predictive modeling scenario where a class label is predicted for a given data which is the “input data.” Classification is used to predict a discrete class label. Given a set of emails classified as spam or not spam or given a handwritten alphabet classified as known character or unknown character from a data set of characters are examples of classification. Regression is a method to identify the relationship between an independent variable and a dependent variable. This is generally used to predict the continuous values of a variable (Kumari and Yadav, 2018). In this study, the authors use some efficient ML techniques including ridge classifier cross-validation, logistic regression, decision tree, k-nearest neighbor classifier (KNN), support vector classifier (SVC), linear support vector classifier (linear SVC), bagging classifier, random forest classifier, Bernoulli naive Bayes classifier (NBC), Gaussian NBC, AdaBoost classifier, gradient boosting classifier and extra tree classifier to predict employability, chance for a high salary and the opportunity to get a pre-placement offer (job offer based on performance in an internship).

Ridge regression is a type of linear regression which is used in a scenario where there is a need to estimate the coefficients, when the independent variables are highly correlated. In other words ridge regression is used to solve multicollinearity issues caused by correlated variables (McDonald, 2009). It is a model that shrinks the coefficients of correlated variables. Ridge regressor computes the optimal ridge regression model based on cross-validation. Ridge classifier cross-validation is an advanced method in which first the target values are converted into {−1,1} and then the data is subjected to regression. Also, cross-validation is used for validating the performance of the model. A supervised learning algorithm can be used to predict the probability of a variable is logistics regression. When the target variable is dichotomous like High/Low or Yes/No or Possible/Not Possible then, logistic regression model helps in prediction. Hence, logistic regression models are used to predict the probability of a discrete outcome (Peng et al., 2002).

A powerful supervised ML technique is decision trees. In a decision tree algorithm, one has to give the details of the input and corresponding output. This has to be done while training the data set. The output data keep splitting according to specific parameters. Decision tree has two prime entities namely, decision nodes and leaves. This powerful ML model explicitly represents decision-making and decisions. It is used for both classification and regression.
KNN classifier is a simple yet efficient non-parametric method for classification (Zareapoor and Shamsolmoali, 2015). A KNN classifier stores all the input data and then classifies and creates a new data point. This new data point is created based on the similarity that exists in the input and its relation to the output. This is how the data set is trained. When the real-time data set is fed after training, KNN classifies it and hence predicts the output using the KNN algorithm that is trained using the test data set. KNN is a well-known statistical technique for pattern recognition and prediction. Support vector machine (SVM) is supervised ML technique. This can be used for both classification and regression scenarios. SVM transforms or alters the data and then based on the transformations the algorithm finds an optimal boundary among the possible outputs. SVC is a clustering algorithm which is non-parametric. It does not build any assumption about the shape or number of clusters in the input data. It is best suited for low-dimensional data. A linear SVC fits the data and gives a best-fit hyperplane which divides the data. Using the hyperplane, one can input certain specific features into the classifier to understand how the predicted output or class is. Both SVC and linear SVC use hinge loss, which is a loss function. This loss function is used for training classifiers. SVC tries to minimize regular hinge loss, whereas linear SVC tries to minimize squared hinge loss (Bansal and Singhrova, 2021). Bagging classifier uses ensemble modeling. In ensemble modeling, many analytical models that are different but inherently related are run and the results are synthesized into a single value. This helps in improving the accuracy of the predictive analysis mode. Using ensemble meta-estimator modeling, bagging classifier places base classifiers on random subsets of the original dataset. Bagging classifiers then aggregate the individual predictions to create a final prediction. Bagging classifier helps in improving stability and accuracy of prediction (Zareapoor and Shamsolmoali, 2015). Random forests or random decision forests also used uses ensemble modeling. It can be used for both classification and regression. A random forest functions by constructing a multitude of decision trees. A swarm of decision trees are created during training the data set. During classification, the output of random forest is chosen as the class selected by most trees. Traditional ML models fail in handling complex and noisy data. Ensemble models perform better in such cases. When an ensemble model is used, it extracts all features and then builds various weak learners using multiple ML techniques. Further to this, a voting system is used to produce better predictive performance (Dong et al., 2020).

Although both bagging classifier and random forests use ensemble modeling, in random forest only a sub-set of features are selected at random from the total. The choicest split feature from the sub-set is taken to split every node in a tree, whereas in bagging classifier all the features are taken into consideration for splitting the node. NB is a probabilistic technique used for prediction and NBC uses Bayes theorem. NBC predicts membership probabilities for every given class so that the probability that a given input data fit to a specific class is predicted. The class that has the maximum probability value is taken as the most likely class. Bernoulli NB is a variant of NB that is a probabilistic technique and uses the Bernoulli distribution Gaussian NB that is another variant of NB which uses the Gaussian distribution. Both are generally used for classification (Zambang et al., 2021). An AdaBoost classifier starts with placing a classifier on the original dataset and then matches extra copies of the classifier on the same dataset where the weights of wrongly classified instances are adjusted so that the succeeding classifiers concentrate more on difficult cases. The AdaBoost method can enhance the efficiency of any ML algorithm. It gives good results with weak learning data sets. Gradient boosting is another ML algorithm that works well with both regression and classification. It creates a prediction model in the form of an ensemble of weak prediction models, like a decision tree. Gradient boosting classifiers are like a collection of decision trees. Random forest builds multiple trees independently, whereas the Gradient boosting method generates stage-wise additive trees (Selvapandian and Manivannan, 2018). Extra trees classifier, which is also a ML algorithm, works by generating a large number of unpruned
decision trees using the training data set. Extra trees classifier has a collection of trees and adds randomization and optimization while building trees (Shtar et al., 2021). The predictions are made by taking the average of the predictions made in decision trees in the regression mode and using the majority voting if run in the classification mode.

2.4 Signaling theory and the abilities of an individual for employability

The signaling theory was originally created by Spence (1973), which was built on observed knowledge gaps between companies who hire individuals for their organizations and the individuals themselves. This model was adapted (Connelly et al., 2011) by several domains like human resource management, business management etc. As per Spence’s job-market signaling model, individuals seeking employment communicate their abilities to the employer (send signal) by acquiring educational qualifications. As per this theory, the employer believes the educational credential to be associated with having the greater abilities and hence have higher chances of being employed. This paper draws its theoretical foundation from signaling theory. As this theory claims that educational qualification increases the ability of an individual, which influences employability and salary, the level of education that an individual acquires is a testimony of the individual’s ability that could be left unobserved by the employer.

2.5 Gap in signaling theory

Signaling theory in education indicates that the level of productivity of an individual is independent of education but the educational qualification acts as a testimony for higher ability. However, this theory does not explain the reason for the mismatch between “education and work” or “education and the disparity in salary” between individuals who earn differently but have the same qualification. Way back in 1970, only 1 to 2% of taxi drivers or truck cleaners or even firefighters were qualified with a college degree; in contradiction to this, today nearly 15 to 20% who take up such jobs have college credentials (Vedder et al., 2013). This aspect is also in contradiction to the Bayesian updating process, which is an assumption in most signaling literature. According to the Bayesian updating process, each new bit of information affects the original probability, and when applied to education, each new and higher educational qualification adds to the individual’s ability and hence improves the chance for a better job and better pay prospects.

Based on this, the objectives of the study have been framed as follows:

1. Use predictive modeling, using ML algorithms, and identify the key aspects that influence the chance of getting a job immediately after pursuing an MBA;

2. Use predictive modeling, using ML algorithms, and identify the key aspects that influence the chance of bagging a high salary (pay package) after pursuing an MBA and

3. Use predictive modeling, using ML algorithms, and identify the key aspects that influence the chance of bagging a pre-placement job offer (job offered based on performance in internship) after pursuing an MBA.

2.6 Choice of attributes

When corporations hire management graduates, they are interested in talented graduates who are problem solvers. The skills, knowledge and abilities that they need has always been changing and is always a mixed bag requirement. This has been the case right from 1990 till date (Grimbly, 1993; Linder and Smith, 1992). The literature indicates that oral communication, interpersonal skills, leadership skills, written communication ability,
decision-making ability, analytical skills, ability to work in a team, work experience, financial skills, technical skills, domain skills, scholastic achievement, internship experience, extracurricular activities, work ethic, nature of past work experience, knowledge of business trends and leadership abilities are all the attributes that influence the decision related to hiring a management graduate (Ornellas, 2019; McCabe, 2010; Lowden et al., 2011; Little, 2001; Yorke and Knight, 2004; Dearing, 2016). The existing research literature also indicates that knowledge on social media and being active on social media also has a significant impact on the chance to be hired (McArhur, 2017). Based on analysis and also interaction with Human Resources (HR) managers who visit B-Schools for recruitment, the attributes chosen for the study include educational background at the graduation, score secured in class 10, score secured in class 12, score secured in graduation, competency in soft skills, participation in co-curricular activities and social engagement status.

3. Research design
The data for the study were taken from a reputed B-School from western India. A convenience sampling was used and the details of 1,202 MBA students over a period of six years were considered for the study. While there are several attributes that impact placements, a set of attributes was selected based on the information sought by the recruiters who ask specific information from the candidates. The recruiting managers used this information to shortlist students for their organizations. All the information that were sought by the recruiters were collected and a comprehensive list of 24 attributes were arrived that were covering information sought.

In total, 24 attributes including educational background at the graduate level (BE, B Tech, B Com, BSc, BBA and others), score secured in class 10 (high, medium and low), score secured in class 12 (high, medium and low), score secured in graduation (high, medium and low), competency in soft skills (high, medium and low), participation in co-curricular activities (high, medium and low) and social engagement status (high, medium and low). The list of attributes is given in Table 2.

The level for each attribute was also defined. Score secured in class 10, class 12 and graduation was defined as “High” if the score was above 80% and was defined as “Medium” if the score was between 60 and 80% and was defined as “Low” if the score was below 60%. Competency in soft skills was rated as high, medium and low based on the group interaction score that was obtained by the student, which was evaluated by experts. This score was stored for all the students over the period of time. B-School gives opportunities to participate in co-curricular activities. The students were given scores of high, medium and low based on the number of co-curricular activities in which they participated. If a student participated in more than ten co-curricular activities, he/she was given a high rating; if it was between 6 and 10, he/she was given a medium rating and any value below 6 was given a low rating. Social engagement status was also at three levels. It was a “High” rating if the student was active on social platforms by posting, at least, one tweet on twitter or one post on LinkedIn – once a week. The rating was medium if it was once in a fortnight and others rated “Low.”

The predictors were taken for the training data set based on the data available. The student placement record of 1,202 gave the information if they got a pre-placement job offer (job offer based on their internship). If they got the offer then it was taken as “Yes” and all others were designated as “No.” The salary was taken as “High” if the annual package was more than the fee paid for the program and was otherwise taken “Low.” The premier B-Schools in India have a structured placement process where the students are given an opportunity to face the corporate who offer jobs to graduating students. The corporates are invited on a daily basis. The first day is termed “Day 0” and the next day as “Day 1” and so on. All the students who were placed within three days of the start of the placement process were
given “High” for employability and all others were given “Low.” Predictive modeling was run to find the attributes with influence and help predicting employability, predicting pre-placement offers and salaries.

4. Data analysis, results and interpretation

Analysis was carried out for predicting the employability of students. Various ML models including ensembles were used in building the prediction model. The steps were executed for data analysis using python and Jupiter notebook as follows:

1. Data preparation: The data were prepared before analysis and modeling. The data were checked for duplication, miscoding, correctness and completeness; the preprocessing of data were completed. Various steps in this stage were selection, cleaning, transformation, handling an imbalanced dataset, feature extraction and new variables (Alyahyan and Duštegor, 2020).

2. Choose candidate ML algorithms: In this step, the authors shortlisted 13 efficient supervised learning algorithms from the literature for classification. The supervised learning approach trains the model with available input and output data, further predicting future data output.

3. Build ML predictive models: The authors used a 70:30 ratio for training and testing datasets and applied cross-validation. In this study, the authors conducted three experiments for building models for predicting employability, the chance for PPO and salary. The output class was binary and following classes were considered for output: for the first model it was YES/NO, for second model also it was YES/NO and for third model its was HIGH/LOW.
Predict using models and evaluate performance: Selecting the best model is a challenge. The authors used selected models to predict employability, the chance for pre-placement offer (PPO) and salary. Accuracy, sensitivity, specificity, precision, area under receiver operating characteristic (ROC) curve area under the curve (AUC) and F1 score were used for measuring performance.

The best model was selected. Figure 2 shows the various metrics used for comparing these classifiers. These models predict “employability” to be high or low. All the algorithms listed in Figure 1 were used to build the model.

It was found that ensemble models were giving better performance than other classifiers such as Gaussian or regression. Gradient boosting classifier was selected as the final model as the accuracy and other performance metrics were better. The model obtained 95.84% accuracy, precision 0.9554 and AUC was 0.9574, which is the metric that helps in evaluating how well a model classifies positive and negative outcomes at all possible cutoffs. Boosting models have less bias, whereas bagging models focus on lowering variance. Also, feature importance was derived from this model, which explains the strength of the prediction. The model unveils that high involvement in co-curricular activities and good soft skills positively influence employability, whereas low soft skills and co-curricular activity negatively impact employability. The features used in building the ML models are listed in Figure 3 and corresponding features are plotted in the x-axis along with the importance on the y-axis.

Predictive analysis was further used to predict PPO to a student. Various ML models including ensembles were used in building the prediction model. Figure 4 shows the various metrics used for comparing these classifiers. The model predicts if a PPO will/will not be offered. All the algorithms listed in Figure 2 were used to build the model.

It was found that ensemble models were giving better performance than other classifiers such as Gaussian or regression. Extra trees classifier was selected as the final model due to accuracy and other performance metrics. The model obtained 96.12% accuracy, precision 0.9634 and an AUC of 0.9605. Extra trees models are faster and computationally cost-effective and hence more robust, whereas bagging models focus more on lower variance. Feature importance was derived from a model enabling explaining ability of the prediction. The model unveils that poor soft skills and low involvement in co-curricular activities lead to no
PPO offer. It was also observed that good soft skills and high involvement in co-curricular activities positively influence PPO offer. Figure 5 shows the feature importance.

Predictive analysis to predict the salary range was the last phase of analysis. Various ML models including ensembles were used in building the prediction model. Figure 6 shows the various metrics used for comparing these classifiers. The model predicts the salary package offered to be high or low. All the algorithms listed in Figure 2 were used to build the model.

It was found that ensemble models were giving better performance than other classifiers such as Gaussian or regression. Extra trees classifier was selected as the final model due to accuracy and other performance metrics. The model obtained 90.30% accuracy, precision 0.9310 and an AUC of 0.8970. Feature importance analysis shows that high co-curricular involvement, good soft skills and high marks in 12th positively impact the salary package offered, whereas low soft skills lead to poor packages. Figure 7 shows the feature importance.
5. Discussion

In this paper, the authors have evidence to contradict the signaling theory in education to prove that educational qualification alone cannot testify to the higher ability for employability and salary. Using predictive analytics, the authors have found that better performance at a lower level of qualification (class 12) influences the ability to earn a higher salary.

In the last decade, employability prediction studies were conducted in various countries. The focus of these studies was to predict employability at graduation, management and technical education level (Ayoubi et al., 2017; Benson et al., 2014; Mezhoudi et al., 2021). However, there is no theoretical evidence found to substantiate the empirical results. In this study, efforts were put to see the significance of signaling theory.

This study used discrete data rather than continuous data to improve predictability. This improvisation also helps remove the bias in the marking schemes of different universities to a certain extent. The data collection and preparation phase has various challenges concerning
the integrity of data. In present study, the authors used the data directly from B-School and hence argue that integrity is maintained. The authors thoughtfully carried out data transformation and feature selection based on descriptive analysis of the raw data. The study also ensured that a balanced dataset is chosen for the model-building to avoid imbalance dataset prediction biases and errors.

The selection of features is very crucial for predictive modeling. Bharambe et al. (2017) used hard skills and soft skills as features, whereas Existing literature adds demographics, employment history and no. of jobs applied as features. In this study, the feature set is slightly different and the authors have considered academic performance, participation in co-curricular activities, social media engagement and communication skills. Though academic performance can be considered similar to hard skills, analysis revealed that graduation-level performance does not impact employability. However, it was found that the performance in class 12 has a greater influence on students’ academic performance and leads to better salary packages. It was also interesting to note that information technology and related studies revealed that hard skills are a significant feature.

The predictive analysis results revealed that ML algorithms can be used to build a predictive model. The literature shows that random forest (Bharambe et al., 2017), decision tree (Piad et al., 2016; Othman et al., 2018) and SVM (Alghamlas and Alabduljabbar, 2019) gave better results in predicting employability. In this study, the results show that ensemble models have better performance outcomes than generic ML classifiers. It was interesting to note that bagging, boosting and extra tress gave an almost equal performance in all three predictive models. In employability prediction the authors recommend gradient boosting classifier, as boosting models have less bias compared to bagging models and give an accuracy of 95%. In the case of PPO and salary prediction the authors selected extra trees, as the model building cost is comparatively lesser here while providing a good performance metrics.

The performance measures selected have some commonality for studies available in literature. Bharambe et al., (2017) used accuracy and time to build the model as performance metrics, while accuracy, F1-score, precision and recall are considered in other studies. In this study, authors selected the relevant performance metrics set such as accuracy and precision for final model selection and observed better results.
The analysis shows that soft skills and participation in co-curricular activities are the major deciding factors to predict employability and a pre-placement job opportunity at the post-graduate management education level. The marks secured in class 12 are one attribute added to this list to predict salary.

6. Conclusion with implications
In this paper, the authors have evidence to contradict signaling theory in education to prove that educational qualification alone cannot act as a testimony for higher ability for employability and salary. The predictive analysis shows that at the post-graduate management education level, soft skills and participation in co-curricular activities are the major deciding factors to predict employability and pre-placement job opportunity, and marks secured in class 12 is one more factor that gets added to this list to predict salary.

The analysis and the predictive model may apply to those Indian B-Schools wherein the quality of students is almost the same. The finding of this study can immensely help management graduates to focus on key aspects that can help them hone their appropriate skills and also can help management institutions to select the right kind of students for their management programs and is a significant contribution of this study. An interesting aspect of this study is that data considered for this study is from a premier Indian B-School. These students who join this B-School go through a rigorous selection process for admissions. The final list of students who join this B-School are in the top three percentile. After a stiff competition from hundred thousand students only a few hundred qualify to join the B-School. Hence, this analysis and model may apply to such B-Schools wherein the quality of students is almost the same and at higher levels. Further, the study can be taken up with Tier-2 and Tier-3 B-Schools to check if these predictions would be good for the students pursuing their post-graduate management program.

References


**Further reading**


**Corresponding author**

Ramakrishnan Raman can be contacted at: director@sibmpune.edu.in

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com