Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 8, 9122-9130

Application Of Decision Tree Classifier To Determine The Selection Of Specialization By Students During Their MBA

Dr. Pranjal Muley[1],Dr.Brijesh Sharma[2], Dr. Dinesh Gabhane[3], Sampada Kulkarni[4], Bhagyashree Khabale[5]

[1]*Associate Professor, VES Business School, Chembur*
[2]*Associate Professor, VES Business School, Chembur*
[3]*AssistantProfessor, Rajeev Gandhi College of Management Studies, Navi Mumbai*
[4]*Student, VES Business School, Chembur*
[5]*Student, VES Business School, Chembur*

Abstract

Master degree in Business management after engineering has now a days become a standard de facto. More and more engineering students prefer to opt for Post graduate in Management. The study is based on supervised learning approach and conducted by using decision tree classifier. Tree based approaches allow predictive models with high precision, reliability and interpretative ease. The study attempt to create tree-like graph or decision-model by classifying the attributes and learn the decisions taken by students for opting either MBA with standard specialization or MBA with "Business Analytics/ Data Science" specialization. Standard specialization is considered as "Non-technical" and Business Analytics/ Data Science specialization is considered as "Technical" in the present article. This model is based on the features that hold the information regarding students such as their age, gender, stream opted in graduation, and final graduation percentage to determine their preference while predicting their respective MBA specialization Data is collected through convenience sampling approach via google survey form.

Keywords Technical, Simple Decision Tree, R Studio, Confusion matrix, ROC curve

1. Introduction

A career in management has happened to prime significance throughout these most recent couple of years. Management is a very large discipline itself in the form of various specialization. Therefore, students who choose Management as a profession move, they have to confront with a question as to which main specialization should be selected. This choice has been influenced by various factors, such as stereotypes for occupations, an individual's skills and abilities, the job market for a particular specialization, the influence of family members, the personality of individuals and many others (Kothari T P et al, 2015). Recent trends show that many engineers, after their B.E. or B.Tech. degree, prefer to opt MBA with various specializations.Fisher (1994), in one of his studies, proposed that the program of MBA should be shifted from delivering knowledge to applying knowledge in real-world situations.

Although many business schools have begun providing their MBA students with business analytics programs and courses. However, the critical function of university ranking is to assess the effectiveness of BAB in producing work market gains for students (Ghasemaghaei et al, 2019).

According to the E.Y. report 2019, the way companies plan to use analytics to drive broad-level decision-making has increased by 50 percent. Similarly, a report by Accenture highlights the importance of Analytics in consumer-packaged-goods (CPG) and the process that companies should develop as an enterprise-wide analytics strategy to harness the power of analytics (Dhariyal L,2020).

The present article attempts to learn the likelihood to opt specialization by students in MBA as Technical or Non-technical. Technical specialization includes "Business Analytics" or "Data Science" whereas Non-technical specialization includes standard specialization namely "Marketing", Finance", "Human Resource" and Operations".

2. Related Work

Data mining applications have influenced almost every field viz Retail, Transports, Media and Healthcare. For instance, Diabetes is a disease that may contribute to various other chronicle diseases in the human beings when affected with it. Gaganjot Kaur et al (2014) have used modified J-48 algorithm in order to mine the data efficiently in case of critical medical records of the people with or without diabetes. They have aimed at increasing accuracy level of the data mining procedure. The study made use of data mining tool WEKA as an API of MATLAB for the generation of modified J-48 classifiers.The results proved that proposed algorithm could attain accuracy up to 99.87%.

Data Science and Machine learning are slowly becoming a part of mainstream Academics and also being widely popular in recent times. Murali, S et.al (2019) says that MBA has become a necessity for top positions of organization. It is the most significant qualification of today's corporate world. Many researches have been conducted to study the factors which are considered important by an individual while deciding upon their career. These factors include the personality types of individuals (Astin, 1993; Pike, 2006; Smart, Feldman &Ethington, 2000), their interests (Kramer, Higley& Olsen, 1994), race/gender, ethnicity, political views (Porter &Umbach, 2006), person–environment congruence (Feldman,Smart &Ethington, 1999; Smart et al., 2000), self-efficacy (Bandura, 1986, 1997), belief (Eccles, 1987) and various others.

AlaaKhalaf H. et al (2018) proposed a model using three built classifiersJ48, Random Tree and REPT Tree to predict Student Performance in Higher Education Institutions. The Weka 3.8 tool was used to construct this model. In another similar research, Bayesian classification method was also used by Bhardwaj K. et al (2011) on student database in order to predict the students' grades basedon previous year performance. The researchers concluded that the study would assist students and teachers in improving student grades. It also helps in identifying those students who need special attention for reducing failing ratio and taking appropriate action in time.

Data mining methods and techniques are used to analyse data at educational institutions in the process of educational data mining. Mashael A. et al (2016) used education data mining techniques especially decision tree to predict students' final GPA based on their courses of previous years. Data mining is largely related to knowledge management as knowledge management includes process of data usage and data mining is the process of data to convert in knowledge. Marjan Laal (2011) recommended in one of his research articles that Colleges and universities should use knowledge management techniques to support any aspect of their mission.

The explosion of educational data is one of the most significant problems that educational institutions are facing. Using this data to improve the quality of managerial decisions in every aspect is a real challenge. Mohammed I.Al-Twijri et al (2015) recommended to use data mining techniques with the help of proposing a data model to extract knowledge from these large datasets. He argued that this model would help in decision making process at strategic levels of higher institutions and also regulates the disciplines of students' admission.

3. Research Gap and Objective

Most of the research work done in higher education domain and reviewed was discussing and arguing Data mining methods were to determine students' performance. The main objective of this study was to determine the likelihood of opting the specialization by First year MBA students in terms of Technical and Non-Technical.

4. Research Methodology

4.1 The sampling:

The sampling frame for the study consisted of the first-year students of Management students of Mumbai region. The population of the study consisted of the first year Management students who had done their graduation in any domain.

4.2 Research Design and Tool:

The study is exploratory in nature. The study used convenience sampling. Data was collected through questionnaire. 150 responses had been received.

The study employed machine learning technique with Decision-Tree classifier to create classification model in order to create a classifier model using existing data and predict the likelihood of opting specialization of Management students. R studio platform was used to execute simple decision tree classifier on the data.

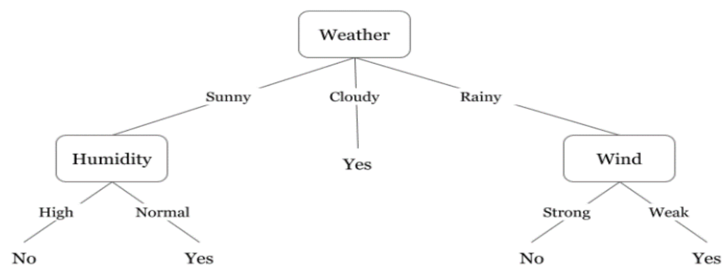
4.3 Classifier selection:

There are many classifier algorithms available for classifying data of binary nature. The decision tree classifier has been selected for this exploratory study because in contrast to other algorithms, decision trees need less effort during pre-processing for data preparation. A decision tree does not need data to be normalized as well as data scaling. The process of building the decision tree is also not significantly affected by missing values in the data. A model made by decision trees is very intuitive and easy to communicate both to technical teams and stakeholders.

4.4 Decision Tree Classifier:

The Decision Tree Classifier is defined by the successive classification of an unknown sample into a class using one or more decision functions (Swain P. H. et al, 1977). In general terms, there is a root node, a number of internal nodes and a number of terminal nodes in a decision tree. The root node and interior nodes, collectively considered as non-terminal nodes. These nodes are linked to each other while forming various decision stages and linked eventually to the terminal nodes representing final classification (Swain P. H. et al, 1977). Decision tree classifiers (DTC's) are used effectively in many fields, including classification of radar signals, character recognition, remote sensing, medical diagnosis, professional systems, and speech recognition, to name only a few. The most significant aspect behind the popularity of Decision tree classifier is assumed to the ability to break down a complicated decision-making process into a series of simpler decisions, offering a solution that is much easier to grasp (Safavian S. R. et al, 1991).

Figure-1: Sample decision tree classifier



(Source: https://www.hackerearth.com/)

5. Results and Discussion

Before applying decision tree classifier on data to build classifier, data was first pre-processed mainly for variables. They were categorical and numeric datatypes. No missing values were found in data. All variables uniformly were converted in numeric data format. The data file is comprised of total 150 observations/ instances that is equivalent to 150 samples in statistics.

There were total 20variables in the file. Fivevariables had been selected considering their relative importance to be used as input in decision tree classifier.The data format of the decision tree classifier is as follows:

Figure-2: Snapshot of data format used in decision tree classifier

Age_group	Gender	stream	Graduation_percentage	Graduation_Degree_Category	P.G_degree_category
20 -25	F	Engineering	70	Engg T	1
26 -30	F	Engineering	99	Engg T	1
>36	M	Engineering	60	Engg N-T	1
20 -25	M	Engineering	7.2	Engg T	0
20 -25	M	Management	70	Others N-T	0
20 -25	F	Commerce	72.8	Others N-T	0
20 -25	F	Arts	77	Others N-T	0
20 -25	F	Commerce	70	Others N-T	0
20 -25	M	Commerce	65	Others N-T	0
26 -30	F	Science	61	Others N-T	0
20 -25	F	Commerce	78	Others N-T	0
26 -30	F	Engineering	70.11	Engg N-T	0
20 -25	F	Commerce	69	Others N-T	0

(Source: Primary Data Source)

One record or instance of the data file is equivalent to one sample in the statistics. A snapshot of the samples in displayed in the **Figure-2**. There weretotal 150records in the data file. The vertical line is known as attribute which is equivalent to a variable in statistics. **Table-1**represents the Graduation degree categories considered in this research. Thereis total 5 attributes or variable namely Age_group, Gender, stream, Graduation_percetange, Graduation_Degree_CategoryandP.G_degree_ category. The last attribute is label attribute represent the entire instance. This label attribute is called classification label or class label. This research has two class labels namely “0” which represents MBA with nontechnical specialization and “1” which represents MBA with technical specialization (**Table-2**).

Stream	Program	Category
Art	BA, BMM	Other NT
Commerce	BCom, BBA	Other NT
Management	BMS	Other NT
Science	B. Pharma, B.SC, B.SC Biotech,	Other NT
Science	B.SC IT, B.Sc. Statistic	Other T
Engineering	Mechanical, Civil, Electrical, Chemical, Production	Engg NT
Engineering	EXTC, IT, Electronics, Computer science	Engg T

Table-1: Graduation degree categories

Program	Specialization	Consideration	Label
MBA/PGDM	Marketing / HR / Operations / Finance	Non-Technical	0
MBA/PGDM	Business Analytics / AI & ML / Data science / Data Analytics / Big data	Technical	1

(Source: Primary Data Source)

Table-2: Post-graduation degree categories
(Source: Primary Data Source)

Dataset or records were divided in training and testing dataset in order to build a decision tree classifier using machine learning approach. Percentage method was used to split records in training and testing datasets. 80% data was used to train the classifier model wherein 20% data was used to test the validity of the model.
Using R studio software application, decision tree classifier was applied on the training data. The result was obtained in the form of Decision tree model and this model is evaluated by confusion matrix and ROC curve measures.
The Confusion Metrix for categorizing students of MBA/ PGDM opting for technical(Business Analytics / AI& ML/ Data science/ Data Analytics/ Big data) or non-technical (Marketing/ Finance/Operations/Human Resource) specialization,is shown in **Figure-3**. This confusion matrix was obtained from training data.

Table-3: Confusion Matrix (training data)

		Predicated	
		0	1
Actual	0	81	10
	1	7	18

(Source: Primary Data Source)

It is clearly seen in **Table-3** that for MBA with Technical specialization, 18 samples were correctly classified, wherein 7 samples were classified incorrectly. Similarly, 81 samples were correctly classified for MBA with non-technical specialization and 10 samples were incorrectly classified. The classification accuracy in the classifier model by confusion matrix was evaluated as 85%.
The confusion matrix obtained from applying the decision tree model on test data is shown in **Table-4**.

Table-4: Confusion Matrix (test data)

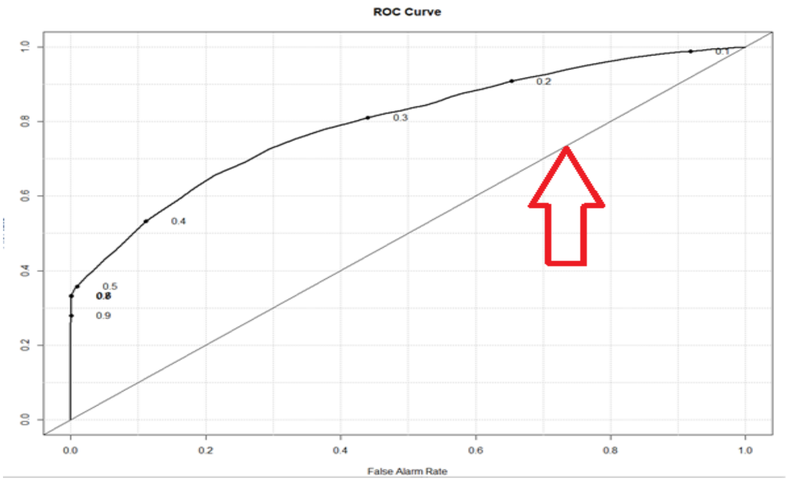
		Predicated	
		0	1
Actual	0	26	3
	1	3	2

(Source: Primary Data Source)

Table-4 indicates that for MBA with Technical specialization, 2 samples were correctly classified, wherein 3 samples were classified incorrectly. Similarly, 26 samples were correctly classified for MBA with non-technical specialization and 3 samples were incorrectly classified. The classification accuracy by confusion matrix was evaluated as 82% which is a good indicator.

Another measure to evaluate classifier model was ROC (Receiver Operating Characteristic curve). It is a graphical representation of the performance of a classification model. It compares the rate at which a classifier makes correct predictions (True Positive) and the rate at which it makes false alarms (False Positive).

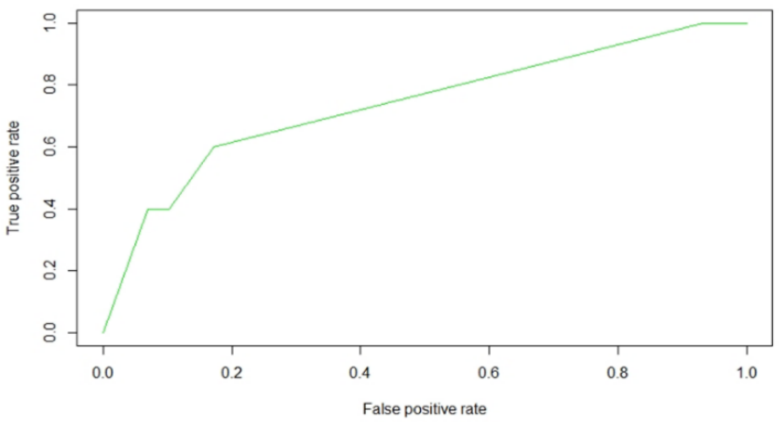
Figure-3: ROC-AUC curve



(Source: Primary Data Source)

ROC is a probability curve and AUC represent the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0s as 0s and 1s as 1s.

Figure-4: ROC-AUC curve to represent accuracy of the model

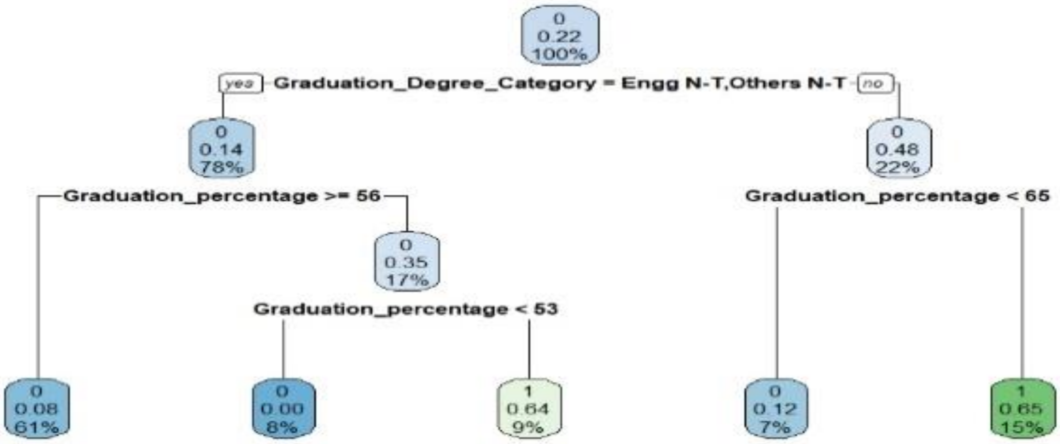


(Source: Primary Data Source)

AUC (Area underthe Curve) value for the Classifier model was evaluated as 73.79%, which determines that there is 74% chance that the model will correctly classify between positive class and negative class. As a result, the model will provide good classification and prediction results for the students in MBA with Technical and Non-technical specialization on the basis of their Graduation degree categories and graduation percentage.

Figure-5 is the decision tree to determine the opting of technical or non-technical specialization by students in MBA. The root node of the tree is Graduation Degree of students.At this stage, data is 100%.

Figure-5: The decision tree model to determine the technical or non-technical specialization in MBA



(Source: Primary Data Source)

The node is further divided if Graduation/Engineering degree is non-technical or technical.In this article Engineering in computer science, IT or Graduation in streams like B.Sc. (IT/Computer Science), BCA are considered Technical degree and Engineering in Instrumentation, Mechanical, Chemical or Graduation in streams like B.Com., BBA are considered non-technical degree. 78% of the total data; considered as a node; was classified as "non-technical" graduation degree whereas 22% data; considered as a node; is classified as "technical" graduation degree.The nodes of 78% and 22% data were further classified on the basis of Graduation Percentage. For the node 78%, if the Graduation Percentage is greater than or equal to 56% then 61% of the 78% node was labelled as "non-technical" specialization in MBA. If the Graduation Percentage is less than 56% then 17% of the 78% node was classified as a separate node, which was further classified. For the node 17%, if Graduation Percentage is less than 53% then 8% of the 17% node was labelled as "Non-technical" specialization in MBA. If Graduation Percentage is between 53% and 56% then 9% of the 17% node was labelled as "Technical" specialization in MBA.

For the 22% node, if Graduation Percentage is less than 65% then 7% of the 22% node was labelled as "Non-technical" specialization in MBA. If Graduation Percentage is greater than 65% then 15% of the 22% node was labelled as "Technical" specialization in MBA.

The overall interpretation from this decision tree classifier is students with technical graduation degree with higher score (above 65%) are more intend to opt MBA with technical specialization (15%). However, there is also a scope among students from non-technical graduation degree with average score (above 53 %) to opt for MBA with technical specialization (9%).

5. Conclusion

This research work attempts to explore and test the decision tree algorithm with the questionnaire on students' data to seek the likelihood of opting MBA with Data science or Business Analytics specialization or other standard specialisations. In this paper, we proposed a data model using simple decision tree of machine learning classification techniques. Data mining algorithms and, in particular, decision tree algorithms can be the best solution for classification and prediction of students' specialisation selection as decision tree algorithms provide high accurate results. They also provide the further guideline to teachers/mentors along with students. Based on the results from the model, it can be said that for data science or business analytics specialization in management, it is not essential that the students should be from technical graduation degree. Students with non-technical graduation degree also opted data science or business analytics specialisation

References

1) AlaaKhalafHamoud, et al., 2018. "Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis". International Journal of Interactive Multimedia and Artificial Intelligence, 5(2):26-31,2018 doi:10.9781/ijimai.2018.02.004
2) Al-Barrak, Mashael A., et al., 2016. "Predicting Students Final GPA Using Decision Trees: A Case Study". International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016. doi: 10.7763/IJIET. 2016.V6.745
3) Al-Twijria, Mohammed I., et al., 2015. "A New Data Mining Model Adopted for Higher Institutions". Procedia Computer Science, Volume 65, Pages 836-844, 2015, doi: 10.1016/j.procs.2015.09.037
4) Astin, A., 1993. "What matters in college: Four critical years revisited". The Journal of Higher Education 22(8), 1993, doi:10.2307/1176821.
5) Bandura, A., 1986. "Social foundations of thought and action: A social cognitive theory". Englewood Cliffs, NJ: Prentice Hall.
6) Bandura, A., 1997. "Self-efficacy: The exercise of control". W H Freeman/Times Books/ Henry Holt & Co.
7) Bhardwaj, K., et al., 2011. "Data Mining: A Prediction for Performance Improvement Using Classification". International Journal of Computer Science and Information Security. Volume 9(4). (2011).
8) Dhariyal, Lekhika, 2020. What is MBA in Business Analytics? Is it a lucrative career option in 2021? Retrieved from https://e-gmat.com/blogs/mba-business-analytics/
9) Eccles, J. S., 1987. "Gender roles and women's achievement-related decisions". Psychology of Women Q. 11, 135–172.
10) Feldman, K. A., et al., 1999. "Major field and person–environment fit: Using Holland's theory to study change and stability of college students". Journal of Higher Education 70: 642–669. doi: 10.1080/00221546.1999.11780802
11) Fisher, M.,1994. "UD MBA plan touts teamwork".The Dayton Daily News, p. 5B, August 9, 1994.
12) Ghasemaghaei, M., et al., 2019. "Impact of MBA Programs' Business Analytics Breadth on Salary and Job Placement: The Role of University Ranking". Communications of the Association for Information Systems,44, pp-pp, doi: 10.17705/1CAIS.04441.
13) Kothari, Tanvi Paras, et al., 2015. "Establishing the Relationship between Marketing Specialization and Personality Traits among Management Students". Paradigm, 19(2) 170–183, 2015.
14) Laal, Marjan, 2011. "Knowledge Management in Higher Education". Procedia Computer Science 3 (2011) 544–549, doi: 10.1016/j.procs.2010.12.090.

15) Pike, 2006, “G.R. Students’ Personality Types, Intended Majors, and College Expectations: Further Evidence Concerning Psychological and Sociological Interpretations of Holland’s Theory”. Research in High Education 47, 801–822 (2006), doi: 10.1007/s11162-006-9016-5.
16) Safavian, S. R., et al., 1991. “A survey of decision tree classifier methodology”. IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660-674, May-June 1991, doi: 10.1109/21.97458.
17) Smart, J. C., et al., 2000. “Academic Disciplines: Holland’s Theory and the Study of College Students and Faculty, Vanderbilt University Press, Nashville, TN”. The Journal of Higher Education, doi: 10.1353/jhe.2005.0017.
18) Swain, P. H., et al., 1977. “The decision tree classifier: Design and potential”. IEEE Transactions on Geoscience Electronics, vol. 15, no. 3, pp. 142-147, July 1977, doi: 10.1109/TGE.1977.6498972.