ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2023.0396

DATA MINING FOR ENHANCEMENT OF GRADUATE ATTRIBUTES

Ritesh Kumar Pandey, Janhavi Obhan and Radhika Kotecha
Department of Information Technology, K. J. Somaiya Institute of Technology, India

Abstract

Analysing students' progress and performance throughout their academic career is critical for boosting their employability. The required skillset and abilities to deal with the ever-changing workplace are increasingly demanded by employers. Graduates must be able to solve problems, communicate well, interact successfully, and think creatively, in addition to possessing good technological talents. Outcome-based education (OBE), which underlines these essential skills, are widely adopted by various educational institutions. Standard assessment measures of OBE have been defined by the Washington Accord as the 12 Graduate Attributes (GA) that can be utilized as relevant benchmarks. Therefore, it is impertinent to formulate an approach which provides a useful system for assessing, projecting, and improving a student's overall academic and extracurricular progress using these Graduate Attributes. The system proposed in this paper applies Data Analytics to predict the progress of the students' skillset and provide them with recommendations to adequately make them the best prospect for any engineering career. Components of the proposed approach have been compared with several baseline approaches and the experimental results demonstrate its efficacy.

Keywords:

Data Analytics, Graduate Attributes, Management Systems, Outcome-based Education, Prediction and Recommendation

1. INTRODUCTION

The world is undergoing several fast shifts, whether in education or business. Companies must improve their competitiveness in this modern context of socio-cultural, economic, and demographic shifts through enhancing human resources. Since the requirements of the fast-paced tech industry are at a superior level in today's day and age, this has rendered the technical talent to stray from being purely basic, as in past. Now, having the basic proficiency with mixture of languages and frameworks which make up the product are nowhere near enough, graduates needs to be equipped with whatever is essential to make the product ready, transcending technical terms. Many individuals lack the qualities to do so.

Therefore, graduate employability has become a problem since there are significant gaps between the graduate abilities obtained at university and the skills required by companies. Although more than 1.5 million engineers graduate each year, a sizable number are inadequately skilled. If the explanation for this skill gap were to be pinpointed, it would be simply a lack of expertise in the appropriate abilities required for growing digital occupations. According to internal research [1] of over 1000 fresh engineers, the most major cause for not being able to get a technological employment was a lack of confidence in applying technical knowledge in real-world circumstances. Even with a plethora of up-skilling and re-skilling choices accessible in our educational system, this specific necessity of assuring deep and powerful experiential learning is weak.

Further, to compete with this expeditious world, the nation needs to strengthen its youth by improving standards of higher education and fortifying skill sets of students. As a result of this globalization, there has been distinct change from education as merely the spread of information to education as the building-blocks of learner competencies, including learning to learn and lifelong learning. What this signifies is that the aim will now need to be shifted to having a deep fundamental understanding as well as upgrading and re-learning skills that are in high demand in the rapidly evolving industry workspace [2]. Outcome-Based Education, which has been widely adopted by various educational organisations, helps in achieving this objective by merging hyper-specialized knowledge with dynamic and cross-sectional capacities, as well as through reinventing curriculums. Hence, there must be provisions and systems developed to record and analyse these outcome-based criteria.

The Washington Accord is an international agreement between bodies responsible for accrediting engineering degree programs on basis of outcome-based education. Since its inception in 1989, over 28 countries worldwide have signed the Accords under their respective bodies to implement the guidelines set forward [3]. These include the National Board of Accreditation in India, amongst several others. The Washington Accord signatories' Graduate Attributes are applicable to the education of professional engineers in all engineering specialties.

The 12 Engineering Graduate Attributes described by the Washington Accords [4] are categorized in Fig.1.

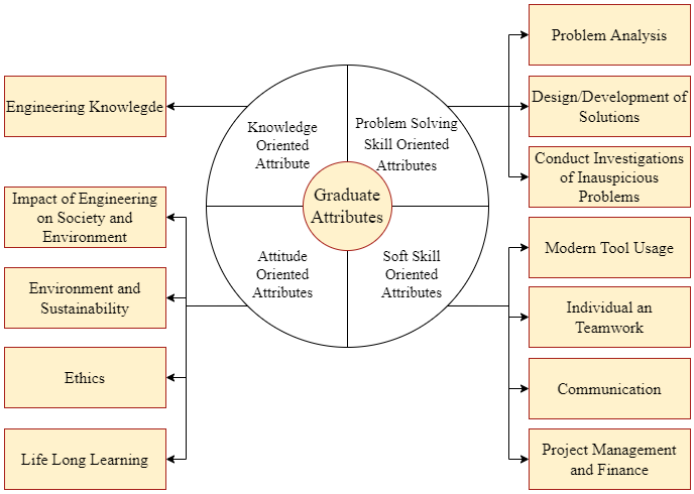

Fig.1. Washington Accord's 12 Graduate Attributes

Graduate Attributes classify what graduates should know, what talents they should have, and what attitudes they should have [4]. They ensure holistic development of a graduate. Graduate Attributes are crucial considerations while developing the curriculum for any university undergraduate programme and are now one of the fundamental sets of higher education outcomes that every graduate should possess [5]. The incorporation of

generic characteristics into the curriculum guarantees that students learn abilities that will better prepare them for the workplace and self-employment. Higher education institutions place an increasing emphasis on establishing graduate qualities in response to requests for accountability and quality assurance systems.

The industry requires two types of traits: technical knowledge and abilities and general attributes, which are covered by the attributes. Generic attributes are soft skills, personal characteristics, and ideals that graduates should develop regardless of their field of study. The regular assessment and analysis of these skills of an individual through the means of these Graduate attributes will highly help in increasing chances of employability. The most important part here is to check for the desirable skills from the very start of higher education to firmly establish such qualities in the Graduates-To-Be.

Data analytics and its branches are rapidly being used across disciplines to develop tools for gaining relevant insights from data and implementing efficient changes. In the sphere of education, [6] for example, such technologies might help with self-regulated learning, student achievement, teacher effectiveness, and institutional decision-making. As a result, the use of data analytics in higher education would enable institutions and instructors to adapt to societal demands and global developments in a timely and efficient manner. Data analysis from several sources throughout an institution would provide a stronger foundation for educational decision-making. As a result, the paper proposes a system that uses Data Analytics to evaluate the Graduate Attributes in students.

1.1 CONTRIBUTION OF THIS WORK

The application area of the work is improvement and automation in education sector. The work researches and proposes a system, which facilitates live progress and enhancement of the skillset of an individual with the following added benefits:

1. The work provides an efficient system to track, monitor, and analyse the performance of undergraduate students through integration and application of data analytics on graduate attributes.
2. Considering the historical and current data regarding the graduate attributes, the proposed system predicts and projects the progress of students over the time, so as to identify areas of improvement by standards of the most sought-after skills in today's world.
3. The work also includes a recommendation system with twofold outcomes. It recommend activities related to the skills deficient in students, and also recommend activities of a students' interest.

This paper is organized as follows: Section 2 presents, for background, a brief description of the previous literature and summarizes the existing work related to the topic. Section 3 highlights the proposed approach for using graduate attributes in skill assessment. Experimental results that demonstrate the performance of the proposed algorithm are stated in Section 4. Section 5 concludes the work and highlights directions for future research.

2. RELATED WORK

The following section explains existing literature and work available on the targeted work. Sub-section 2.1 covers the research available on Graduate Attributes and the relevant existing work. Since the work focuses on prediction and recommendation of enhancing Graduate Attributes, their approaches and adoption for the system have been presented in sub-sections 2.2 and 2.3 respectively.

2.1 LITERATURE REVIEW ON GA ENHANCEMENT

Graduate characteristics are far less generic and far more specific than those imparted by many educational organisations. As Simon C. Barrie [7] argues, this is a more fundamental framework of reference for the competency-based higher education paradigm than what combination of skills, traits, and knowledge should be included on the graduate's shopping list; it is about the nature of the items on the list, as well as the structure of the list itself.

When understood and assessed, graduate qualities can help to improve learning by connecting it to the world of work and allowing our graduates to immerse themselves in global communities. The community can move towards better outcome-based teaching, learning, and assessment model if work is placed into monitoring the growth of graduate qualities in our curricular, co-curricular, and extra-curricular domains.

After introducing graduate attributes at higher education institutions, Jennifer Hill, Helen Walkington, and Derek France suggest that [8] it is necessary to consider how to verify that students are developing certain attributes. Although graduate characteristics are frequently referenced in curriculum documents, their successful integration into developmental processes in the classroom has proven difficult [9].

Students' performance in taught units can be evaluated and graded. Indeed, "explicit embedding in assessment" is the strongest proof of graduate attribute attainment. Moreover, using complementary sources of data, such as student perceptions and extra-curricular participation, to evaluate the accomplishment of graduate qualities is arguably the best approach to do so. Another strategy to assist graduates in managing their own employability is to help them become more aware of the qualities that employers want. If the causes of their mismatch in their views of these graduate traits can be determined, specific efforts can be made to correct them [10].

Finally, Fraser and Thomas [11] discuss that students must actively participate in the building of their own student identities, graduate characteristics, and emergent professional identities, rather than having their identities constructed for them through integrated systems and implementation. Co-curricular activities play a significant role in promoting a more student-centred partnering approach. Students create graduate qualities in this way because they are important to their sense of self, and as a result, they are aware of the talents they have acquired during their studies and can explain them to employers clearly.

The topic of evaluating and analysing GAs, on the other hand, is still an uncharted area. The authors of [12] provide a criteria-based assessment system that allows for an institution-wide

comparison of different GA acquisition levels but fails to analyse it on an individual level. The authors in [13] highlight the purpose of investigating students' progress using performance indicators, which work similarly to graduate attributes. If these estimators of performance are correctly identified, then they can be used to efficiently plan corrective actions to improve the attrition rate and consequently improve placement statistics. In [14] the authors try to assess the knowledge and competency of graduates using the parameters of Graduate Attributes on curricular factors. The lack of an efficient system for evaluation is highlighted across these papers.

2.2 BASELINE PREDICTION ALGORITHMS

A part of analysing and improving the skillset must include the prediction of these skills over a period of time. Many mathematical advances including data analytics [15]-[18] have been made in using algorithmically rich systems to predict numerical data. Many algorithms have been researched and tested, ranging from several types of regressors [19] to neural networks. The description of the few prediction algorithms which are used for the study and how they related to current system are presented below:

- **Lasso Regression:** The Least Absolute Shrinkage and Selection Operator (LASSO) [20] is a Least Square Method modification that works well when the number of features is less than the number of observations. It adopts the L1 norm, which is equivalent to the absolute value of the coefficients' magnitudes.

In context of proposed work, LASSO chooses features, which are GA scores in initial years, and shrinks them by reducing the coefficients of others to zero. Consequently, it predicts the score at graduation by estimating sparse coefficients.

- **Artificial Neural Network:** [21] The activation function of the node in the output layer may be used to reduce neural networks to a classification or regression model. The output node in a regression issue has a linear activation function (or no activation function). A linear function has a continuous output that ranges from -inf to +inf. As a result, the output layer will be a regression-based model that is a linear function of the nodes in the layer preceding the output layer, and the same as been applied to test proposed system. The system flow of ANN Model is shown in Fig.2.

For the system covered by the study, the input variables of GA Score in First, Second and Third Year are *x1*, *x2* and *x3* respectively and *y* is the predicted GA Score at graduation.

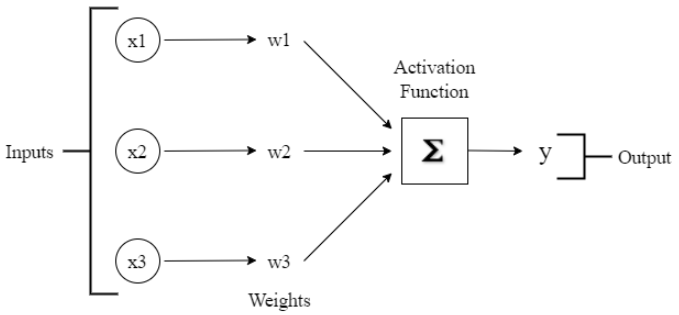

Fig.2. Artificial Neural Network Model

- **Linear Regression:** If a linear connection exists, linear regression [22] uses some independent variable X to forecast the value of a dependent variable *Y*. A straight line can be used to show this connection. The procedure is known as multiple linear regression when there are more than one independent variable. Cost function of multivariate linear regressor used for the current model is given by Eq.(1).

$$Y_i=\beta_0+\beta_1 X_{i1}+ \beta_2 X_{i2}+ \beta_3 X_{i3} \quad (1)$$

For proposed system,

Y_i = Predicted GA Score at Graduation, X_{i1}, X_{i2}, X_{i3}= GA Scores in First, Second and Third Year respectively. β_0 = Intercept and β_1, β_2, β_3 = Regression Coefficients of the respective features X_{i1}, X_{i2}, X_{i3}

- **Random Forest Regression:** A random forest [23] works as a group of binary regression trees. These vast numbers of binary regression trees are generated using an independent subset of variables. The decision trees are built using bootstrapped samples from the dataset, and the variables to divide are chosen at random via Random Forest. The system flow of Random Forest Regressor Model is shown in Fig.3.

In case of the targeted work, it formulates a tree based on the dataset of GA scores over the years and represents the predicted scores on graduation as the leaf nodes of the tree.

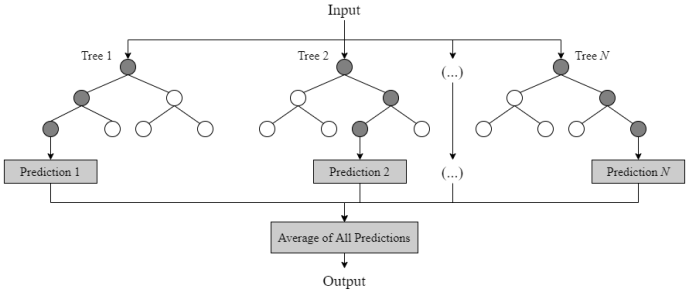

Fig.3. Random Forest Regressor Model

- **Support Vector Regression:** A supervised machine learning algorithm, support vector machine [24], can be used for classification, regression, or outlier identification. SVR's primary concept is to identify the optimum fit line. The best fit line in SVR is the hyperplane with the greatest number of points. The SVR model retains all the algorithm's major characteristics. Because of its cost function, which excludes any training data near to the model prediction, its only reliance is on a training data subset for model development.

For the presently discussed GA scores, which may show variation due to subjectivity of participation, SVR can fit the potentially curved dataset to give accurate predictions.

2.3 BASELINE RECOMMENDATION ALGORITHMS

Historically, collaborative filtering [25] and content-based filtering [26] have been the two major methods for developing recommender systems. Collaborative filtering, in its purest form, depends on the discovery of user community preference patterns. Content-based filtering algorithms, on the other hand, simply take a user's past preferences into consideration and attempt to construct a preference model based on a feature-based

representation of the content of recommended items. While collaborative filtering has a majority in recommendation systems since it considers the quality of the items in the recommendation process, [27] discusses several recent trends in Content-Based Filtering, which accredit much more benefits to the process.

The authors in [28] further discuss the advantages of content-based filtering over collaborative. First off, since suggestions are solely based on user ratings, they are tailored to each user's preferences. In addition, if more users are added to the system, the resources and time required to produce a suggestion do not rise since content-based systems make recommendations for a single user. As a result, the algorithm for providing suggestions can handle many users. Additionally, content-based recommender systems have intrinsic security against fraudulent item generation since they do not rely on user data to make suggestions.

The content-based strategy is appropriate in circumstances or domains where there are more things than consumers [29]. This relates to the current system, wherein the number of events attended and available for suggestions are large in comparison to a single user. As content-based filtering is tailored to each user's preferences, and the recommendations need to be tailored to the previously participated events in proposed system, literary research shows that using this method would be appropriate for the purpose of the proposed system.

Further, checking similarities of keywords of events is one of the ways to go about checking similarity between events. Keywords can be a single word, or many words related to the specifics of the event. Authors in [30] propose a novel particularly useful form of tailored recommendations based on keywords as a scalable and optimized version of user-based recommendation systems. This system is shown to have better accuracy in results or recommendations to the users.

The keywords can consequently be compared with each other using similarity measures. A similarity measure is a function that computes the degree of similarity between terms. [31] explores that with similarity measures, it is possible to sort keywords by importance and consequently the events related to those keywords. Various similarity measures are discussed by the authors like Cosine Similarity, Euclidean Distance, Jaccard Coefficient and Pearson Correlation Coefficient.

It is useful to compare several similarity measures for the precision of recommendation given by the proposed system. Considering that *A* and *B* are two sets or vectors of keywords of previously participated events and upcoming events respectively that need to be compared, for *n* datapoints inside the sets, [32] describes the specifics of similarity measures for the targeted work are explained:

- **Jaccard Distance:** Jaccard distance [33] defined as the cardinality of the common elements of the sample sets divided by the cardinality of the uncommon and unrepeated elements of the sets. Jaccard distance (*J*) is given by Eq.(2).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

- **Euclidean distance:** Euclidean distance [34] is calculated by taking the finding sum of squared differences between corresponding elements of the sample arrays or vectors and consequently square rooting it. Euclidean distance (*E*) is given by Eq.(3).

$$E = |A - B| = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \tag{3}$$

- **Cosine Similarity:** Cosine similarity [35] evaluates the similarity between two arrays or vectors, and it is appropriate since it concentrates on the vector's direction rather than its magnitude. It can be used with TF-IDF. Cosine distance ($cos(\theta)$) is given by Eq.(4).

$$\cos(\theta) = J(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{4}$$

While they adopt a method like the proposed system, none of the existing systems automates the task digitally using AI. Further, the existing systems fail to consider any co-curricular or extracurricular activities undertaken by the student, which contribute greatly to the skills described by graduate attributes. There is also a consequent lacking in the improvement of GA-measured competency, which can be brought about by the efficient implementation of a prediction and recommendation system.

3. PROPOSED APPROACH

The proposed system aims to automate the assessment of the skill-set of a student based on Graduate Attributes and to automate the same using advanced AI algorithms for prediction and recommendations. The formulated approach of the system is represented in Fig.4.

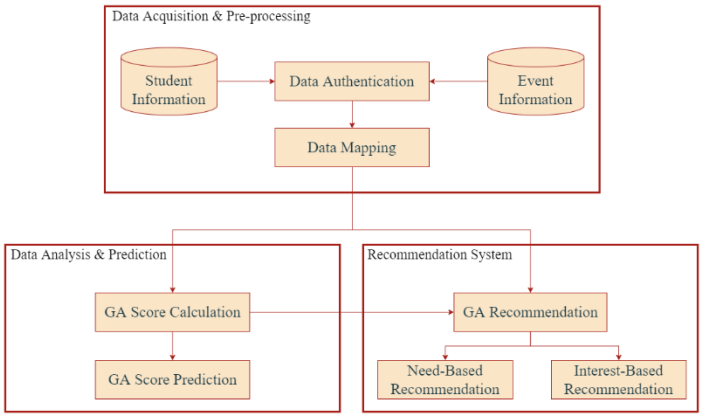

Fig.4. Proposed System

As depicted in Fig.4., the system is divided into three phases each of which is explained in detail as follows:

3.1 DATA ACQUISITION AND PRE-PROCESSING

The dataset used for training, validating, and testing contains details of activities in which a student has taken part along special achievements. The proof of involvement in the activity or event is also stored for authentication purposes. Additionally, it contains the details regarding past and upcoming events, all mapped to each of the individual Graduate Attributes that they inculcate in

the attendees. It also contains the relevant keywords for each event highlighting their focus.

Authentication of the data is done by cross-checking the proof of participation. Once it has been verified, the data is divided into various categories and the respective GAs are mapped and set for further process.

3.2 DATA ANALYSIS AND PREDICTION

The data is processed to allot the necessary GA scores for that participation and the same gets added to the participation history of the student. Basically, the GA scores are calculated based on the event category as well as participation level. The scores are weighted according to the extensivity of the event category as well as the achievement level, i.e., from participation and qualification up to winning titles.

Once the scores have been calculated, the algorithms for predicting the GA score at the end of undergraduate studies are implemented. These consider the scores from the previously recorded data and analyse the current score of the student. The Machine Learning models based on numerical prediction algorithms discussed in the literature implemented on the given dataset.

3.3 RECOMMENDATION SYSTEM

Further, the activities in the participation history of a student and the upcoming activities in the dataset are cross analysed to give two kinds of recommendations as follows:

3.3.1 Need-based Recommendation:

Uses a threshold for each GA (based on batch-wise average generated from data) to recommend activities which map high on the GAs in which the student falls below average.

3.3.2 Interest-based Recommendation:

Uses recommendation algorithms to compare the kind of activities attended by the student in the past to suggest similar upcoming events.

The Fig.5 describes the method used for interest-based recommendations. This kind of recommendation utilizes the keywords from the event information data and calculates the similarity measures of events previously attended by the student with all upcoming events. The upcoming events with the highest similarity to the participation history are recommended accordingly.

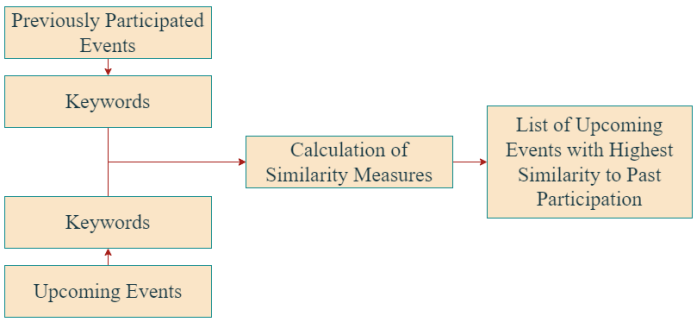

Fig.5. Proposed Model for Interest-Based Recommendations

The similarity measures discussed in sub-section 2.3 of Related Work are implemented to build the interest-based recommendation system.

The effectiveness of proposed approaches has been verified through experimentation and the next section describes the details of the same.

4. IMPLEMENTATION AND RESULTS

4.1 DATASET

The dataset of 4000 students is collected from the records of the engineering undergraduate college where the research is based. It consists of the details of the participation of students and consequently calculated GA scores of over 4 years. The use of a dataset from a live source bolsters the accuracy of the suggested approach in a real-world case. The summarization of the GA scores in the dataset over the four years of education is given in Table.1.

Table.1. Summary of the GA Score Data

GA Score	First Year (x1)	Second Year (x2)	Third Year (x3)	Graduation (y)
Mean	1.1691	2.7797	4.3264	6.9772
Five Number Summary				
Minimum	0.0168	0.9601	2.1030	3.7601
Q1	0.9096	2.1710	3.7903	5.9148
Median	1.2612	2.8561	4.2466	7.0707
Q3	1.4468	3.3453	4.7757	8.0387
Maximum	2.096	4.426	6.7889	10.2162

4.2 IMPLEMENTATION DETAILS

The data shows a uniform increase in GA scores over the years gradually from inception of an undergraduate degree to graduation. The Fig.6 visualizes the trend of the average GA Score over the years.

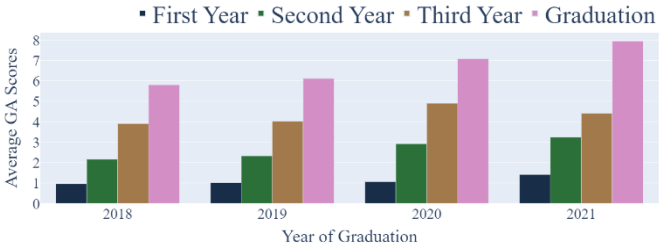

Fig.6. Trends of Average GA Score over the years from dataset

While the GA scores remain mostly linear, there are some variances. This can be accounted for by the subjectivity in the participation of different students throughout the course of the education. Fig.7. visualizes the outliers as exceptional cases in the GA scores.

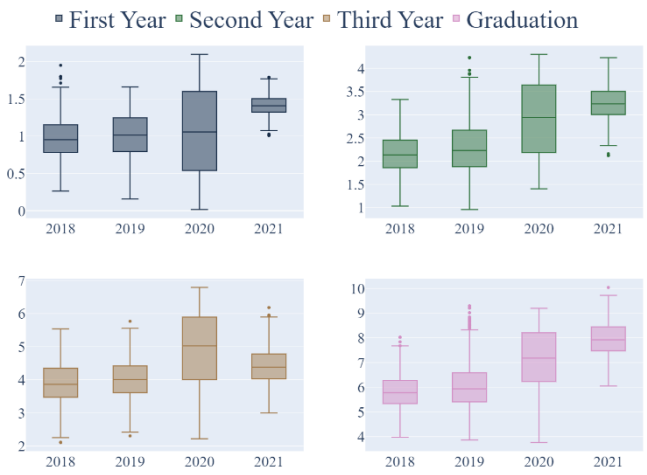

Fig.7. Outlier Pattern Analysis

The variability in the scores of first, second, and third years as correlated to scores on graduation are shown in Fig.8a., Fig.8b., and Fig.8c. respectively.

Fig.8(a). Correlation of GA Score of First Year and Graduation

Fig.8(b). Correlation of GA Score of Second Year and Graduation

Fig.8(c). Correlation of GA Score of Third Year and Graduation

The detailed analysis of the data suggests that while the data follows a gradual increase in the GA score over the year, a prediction model which fits the curve instead of fitting linearity might be optimal for the dataset as it tends to cover exceptional cases too.

The Table.2. describes the Parameters and Key Inputs defined for several prediction models tested.

Table.2. Algorithmic Parameters and Key Inputs

Algorithm	Parameters/Key Inputs	Values
Lasso Regression	Regularization Strength	1.0
	Tolerance for Optimization	0.06
	Constant in Decision Function	6.735
	Maximum number of Iterations	1000
	Feature Selection Method	Cyclic
Artificial Neural Network	Number of Layers	6
	Type of Layers	Dense, Dropout
	Activation Function	Relu, Linear
Linear Regression	Constant in Decision Function	1.672
	Importance Coefficient of Features	[1.030, 0.766, 0.433]
	Normalization of Input	False
Random Forest Regression	Minimal Cost-Complexity Pruning	2.41
	Importance Coefficient of Features	[0.293, 0.489, 0.218]
	Quality Measure Function	Squared Error
	Number of Trees	100
	Randomness of Bootstrapping Method	24
Support Vector Regression	Type of Kernel	rbf
	Constant in Decision Function	6.691
	Number of Support Vectors	749
	Use of Heuristic Shrinking Function	True
	Degree of Polynomial Kernel Function	1
	Size of Kernel Cache	200
	Penalty Rate	0.1
	Tolerance for Stopping Criterion	0.03
	Kernel Coefficient	Scale
Need-Based Recommen-dation	Threshold	Average GA Score
Interest-Based Recommen-dation	Similarity Measures	Jaccard Distance, Euclidean Distance, Cosine Similarity
	Base Input Parameter for Similarity Comparison	Keywords of Events

The proposed model for GA Score Prediction is trained, tested, and validated on the current dataset using several machine

learning algorithms described in sub-section 2.2 of Related Work. These algorithms are evaluated using several Performance Metrics.

Considering parameters as follows, $y \rightarrow$ Value of GA Score, $\bar{y} \rightarrow$ Average value of GA Score, $\hat{y} \rightarrow$ Predicted value of GA Score and $N \rightarrow$ Number of Data Points. The performance metrics are defined as:

- **R^2:** R-squared [36] is a metric in statistics that measures the proportion of the variation explained by an independent variable or variables in a regression model for a dependent variable. The formula for R-squared is given in Eq.(5).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \quad (5)$$

- **MAE:** The Mean Absolute Error [37] gives a measure of the difference in error between two observations expressing the same criteria. The formula for MAE is given in Eq.(6).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \quad (6)$$

- **MSE:** Mean Squared Error [37] estimates the average of the squared errors, or the average squared difference between the predicted and actual values. The Eq.(7) denotes the mathematical formula for MSE.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \quad (7)$$

- **MAPE:** Mean Absolute Percentage Error [37] is a statistical measure of forecasting technique prediction that operates in the same manner as model accuracy. The Eq.(8) and Eq.(9) formulae MAPE and Accuracy respectively.

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{y_i} \quad (8)$$

$$Accuracy = 100 - MAPE \quad (9)$$

4.3 RESULTS AND DISCUSSION

The proposed algorithms have been verified using the performance metrics to compare their efficacy. Table.3. describes the comparison of the algorithms using the performance metrics.

Table.3. Comparison of Performance of Prediction Algorithms

Algorithm	R^2 Score	MAE (in GA Score)	MSE (in GA Score)	Accuracy
Lasso Regression	0.0027	1.0290	1.5282	84.09%
Artificial Neural Network	0.7639	0.3486	0.3598	89.16%
Linear Regression	0.7779	0.3374	0.3384	92.01%
Random Forest Regressor	0.8163	0.2476	0.2798	94.31%
Support Vector Regression	0.9095	0.1085	0.2402	96.93%

The analysis of the dataset suggested that a model that fits a non-linear curve would perform better for the prediction and the lack of the same can account for the inaccuracy in some of the prediction algorithms.

Artificial Neural Network performs better when there is a need to model complex patterns and predictions. The Fig.6 shows that there is a gradual increase in the GA Score over the year, i.e., the GA Score data of students follow a linear relationship and hence, the regression algorithms perform better than the neural network as supported by Table.3.

Generally, all regression models fit a line. Fig.8(a)., Fig.8(b). and Fig.8(c). depict that a model that fits a linear relationship between the data points would be inefficient as it would not learn the exceptional trends and patterns. Like other regression models, the Support Vector Regression Model also fits a line to the given data by minimizing the cost function. However, in SVR, with the help of a non-linear kernel, data can be fitted over a curve rather than a line. By fitting a curve, the inconsistent pattern present in the data for some students is also learned by the SVR model and thus performs better than other regression models. Based on thorough analysis and results showcased in Table.2., it is found that Support Vector Regression is the most optimal algorithm for predicting the GA Scores of the student by the end of their graduation.

Further, recommendation models for both Need-Based and Interest-Based Recommendation Systems were trained and tested out on the responses of the students. In correlation with the collected event database, the average scores attained by the students for all 12 Graduate Attributes are also analysed and represented in Fig.9. The figure also highlights the increment in scores of each attribute upon participating in activities as suggested by Need-Based Recommendation System.

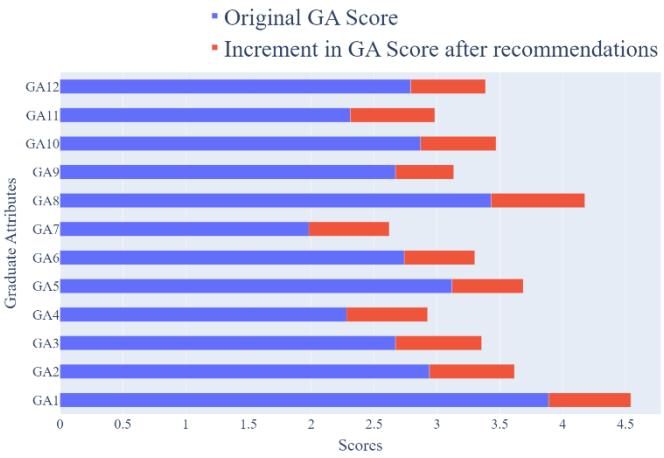

Fig.9. Improvement in GA Score after using proposed system

The suggestions of activities by the Interest-Based Recommendation Model and the future activities attended by students were compared for each of the similarity measures and the results of the same are summarized in Table.4.

Table.4. Comparison of Performance of Similarity Measures

Similarity Measure	Acceptance Rate of Recommendations
Jaccard Distance	85.34%
Euclidean Distance	89.78%
Cosine Similarity	94.52%

The Table.4. denotes that the students tend to participate more in the activities as recommended by the model which uses Cosine Similarity Measure along with CountVectorizer class for event keywords vectorization. This benefits in improved engagement of students in different events. Both Fig.9. and Table.4. indicate that there has been an increase in the participation by the students in different activities as per the recommendations thereby indicating an improvement in GA Score as well as the skill of the individual.

5. CONCLUSION AND FUTURE SCOPE

The application of the research paper links a student's talents learned during their education through numerous activities to their likelihood of possessing a strong skill set. The proposed work builds a system where a person may monitor their growth based on a variety of project contests, competitions for developing interpersonal skills, seminars, webinars, and actively participating in sports and cultural events. The study formulates and implements a three-phase system. First phase being the prediction of GA score upon graduation for which various algorithms were tested. It is found that the Support Vector Regression Model is the most suitable Machine Learning model for predicting the given dataset of GA Scores. Although current systems may analyse a person's skill set adequately, they do not offer suggestions for improvement, which this study incorporates by the means of two different kinds of recommendation systems – Need-Based and Interest-Based, which form the next two phases of the proposal. The suggestions made by the implementation point out the skills that a student needs and offer exercises to improve and strengthen these traits. On experimentation, it is found that the Cosine Similarity Measure performs the best for interest-based recommendations. Overall, the recommendations have a high acceptance rate. The paper achieves its goal of giving an accurate and comprehensive analysis of the student's skill set by investigating how Graduate Attributes may be incorporated into their assessment. The projection of grades upon graduation will also instil in pupils an attitude of continuous skill improvement.

As a future scope, the data can be drilled down to include the curricular learnings of students which can be incorporated in training and testing. Additionally, there are possibilities of mining the data to find patterns in students' individual as well as overall progress and accordingly arrange skill-boosting activities.

REFERENCES

[1] N. Mahadevan, "80% Engineers are Unemployed: How Can we Prepare Engineers for the Jobs of Tomorrow?", Available at: https://www.indiatoday.in/education-today/featurephilia/story/80-engineers-are-unemployed-how-can-we-prepare-engineers-for-the-jobs-of-tomorrow-1468240-2019-03-01, Accessed at 2009.

[2] A. Sahasrabudhe, "Employability of Engineering Graduates", *Proceedings of Workshop on Outcome Based Education and NBA Accreditation*, pp. 1-9, 2014.

[3] International Engineering Alliance, "Washington Accords", Available at: https://www.ieagreements.org/accords/washington/, Accessed at 2022.

[4] International Engineering Alliance, "25 years of the Washington Accords", Available at: https://www.ieagreements.org/assets/Uploads/Documents/History/25YearsWashingtonAccord-A5booklet-FINAL.pdf, Accessed at 2014.

[5] R. Moalosi, M. Tunde Oladiran and J. Uzaik, "Students' Perspective on the Attainment of Graduate Attributes through a Design Project", *Global Journal of Engineering Education*, Vol. 14, No. 1, pp. 1-13, 2012.

[6] A. Nguyen, L. Gardner and D. Sheridan, "Data Analytics in Higher Education: An Integrated View", *Journal of Information Systems Education*, Vol. 31, No. 1, pp. 61-71, 2020.

[7] S. Barrie, "Understanding What We Mean by the Generic Attributes of Graduates", *Higher Educations*, Vol. 51, pp. 215-241, 2006.

[8] J. Hill, H. Walkington and D. France, "Graduate Attributes: Implications for Higher Education Practice and Policy", *Journal of Geography in Higher Education*, Vol. 40, No. 2, pp. 1-12, 2016.

[9] D. Thompson, L. Treleaven, P. Kamvounias, B. Beem and E. Hill, "Integrating Graduate Attributes with Assessment Criteria in Business Education: using an Online Assessment System", *Journal of University Teaching and Learning Practice*, Vol. 5, No. 1, pp. 1-12, 2008.

[10] C. Lee and S. Chin, "Engineering Students' Perceptions of Graduate Attributes: Perspectives from Two Educational Paths", *IEEE Transactions on Professional Communication*, Vol. 54, No. 1, pp. 1-18, 2017.

[11] K. Fraser and T. Thomas, "Challenges of Assuring the Development of Graduate Attributes in a Bachelor of Arts", *Higher Education Research and Development*, Vol. 32, pp. 545-560, 2013.

[12] D. Ipperciel and S. Elatia, "Assessing Graduate Attributes: Building a Criteria-Based Competency Model", *International Journal of Higher Education*, Vol. 13, pp. 23-35, 2014.

[13] K. F. Li, D. Song and F. Rusk, "Predicting Student Academic Performance", *Proceedings of International Conference on Complex, Intelligent, and Software Intensive Systems*, pp. 1-7, 2013.

[14] M. Symes, D. Ranmuthugala, C. Chin and A. Carew, "An Integrated Delivery and Assessment Process to Address the Graduate Attribute Spectrum", *Proceedings of the International Conference on Engineering and Technology Education*, pp. 1-16, 2011.

[15] R. Kotecha and S. Garg, "Preserving Output-Privacy in Data Stream Classification", *Progress in Artificial Intelligence*, Vol. 6, pp. 87-104, 2017.

[16] K. Prakash and K. Selvakumari, "Mathematical Modelling and Big-Data Analytics for Student Performance", *Journal of Physics: Conference Series*, Vol. 1850, pp. 562-574, 2021.

[17] L. Vitoria, M. Ramli, R. Johar and M. Mawarpury, "A Review of Mathematical Modelling in Educational Research in Indonesia", *Journal of Physics: Conference Series*, Vol. 1882, pp. 341-345, 2020.

[18] A. Garg, U.K. Lilhore, P. Ghosh, D. Prasad and S. Simaiya, "Machine Learning-based Model for Prediction of Student's Performance in Higher Education", *Proceedings of*

International Conference on Signal Processing and Integrated Networks, pp. 162-168, 2021.

[19] N. Ashfaq, Z. Nawaz and M. Ilyas, "A Comparative Study of Different Machine Learning Regressors for Stock Market Prediction", *Proceedings of the International Conference on Engineering and Technology Education*, pp. 1-8, 2021.

[20] S. Alsabah, "Estimation Parameters of Lasso and Ridge Regression Models with Application", Master Thesis, Department of Computer Science, University of Kerala, pp. 1-120, 2020.

[21] M. Mishra and M. Srivastava, "A View of Artificial Neural Network", *Proceedings of International Conference on Advances in Engineering and Technology Research*, pp. 1-3, 2014.

[22] K. Kumari and S. Yadav, "Linear Regression Analysis Study", *Journal of the Practice of Cardiovascular Sciences*, Vol. 4, pp. 33-36, 2018.

[23] M. Schonlau and R. Zou. "The Random Forest Algorithm for Statistical Learning", *The Stata Journal*, Vol. 20, No. 1, pp. 3-29, 2020.

[24] F. Zhang and L. O'Donnell, "Support Vector Regression", *Methods and Applications to Brain Disorders*, pp. 123-140, 2020.

[25] P. Singh, P. Pramanik and P. Choudhury, "Collaborative Filtering in Recommender Systems: Technicalities, Challenges, Applications, and Research Trends", *Proceedings of the International Conference on New Age Analytics*, pp. 1-5, 2020.

[26] N. Karbhari and V. Shinde, "Recommendation System using Content Filtering: A Case Study for College Campus Placement", *Proceedings of International Conference on Energy, Communication, Data Analytics and Soft Computing*, pp. 963-965, 2017.

[27] P. Lops, D. Jannach and C. Musto, "Trends in Content-Based Recommendation", *The Journal of Personalization Research*, Vol. 29, pp. 239-249, 2019.

[28] C. Zisopoulos, S. Karagiannidis, G. Demirtsoglou and S. Antaris, "Content-based Recommendation Systems", *Proceedings of the International Conference on Engineering and Technology Education*, pp. 1-5, 2008.

[29] S. Philip, P. Shola and A. John, "Application of Content-Based Approach in Research Paper Recommendation System for a Digital Library", *International Journal of Advanced Computer Science and Applications*, Vol. 5, No. 10, pp. 1-13, 2014.

[30] V. Savadekar and P. Gosavi. "Towards Keyword Based Recommendation System", *International Journal of Science and Research*, Vol. 3, No. 11, pp. 1-15, 2014.

[31] M. Magara, S. Ojo and T. Zuva, "A Comparative Analysis of Text Similarity Measures and Algorithms in Research Paper Recommender Systems", *Proceedings of the Conference on Information Communications Technology and Society*, pp. 1-5, 2018.

[32] O. Zammit, S. Smit, C. Raffaele and M. Petridis, "Exposing Knowledge: Providing a Real-Time View of the Domain Under Study for Students", *Proceedings of International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 122-135, 2019.

[33] J. Hancock, "Jaccard Distance (Jaccard Index, Jaccard Similarity Coefficient)", *Dictionary of Bioinformatics and Computational Biology*, Vol. 12, No. 1, pp. 1-12, 2004.

[34] L. Liberti, C. Lavor and A. Mucherino, "Euclidean Distance Geometry and Applications", *Society for Industrial and Applied Mathematics Review*, Vol. 56, No. 2, pp. 1-16, 2012.

[35] A. Lahitani, A. Permanasari and N. Setiawan, "Cosine Similarity to Determine Similarity Measure: Study Case in Online Essay Assessment", *Proceedings of International Conference on Cyber and IT Service Management*, pp. 1-6, 2016.

[36] J. Han, H. Tong and J. Pei, "*Data Mining - Concepts and Techniques*", University of Illinois Publisher, 2022.

[37] G. James, D. Witten, T. Hastie and R. Tibshirani, "*An Introduction to Statistical Learning - with Applications in R*", Springer, 2013.