Research-to-Practice for Peer-to-Peer Learning in Engineering Education using Ensemble Methods to Deploy a Lifecycle Design Roadmap

Willie L. Brown, Jr.
Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
wlbrown@umes.edu

Ibibia Dabipi
Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
ikdabipi@umes.edu

Dinesh Sharma
Department of Business, Management and Accounting
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
dksharma@umes.edu

Lei Zhang
Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
lzhang@umes.edu

Weiwei Zhu-Stone
Department of Computer Science and Engineering Technology
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
wzhu@umes.edu

Lanju Mei
Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
lmei@umes.edu

Alvernon Walker
Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
awalker1@umes.edu

Tiara T. Cornelius
Department of Mathematics
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
tdturner@umes.edu

Jason Cornelius
Department of Mathematics
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
jmcornelius@umes.edu

Urban Wiggins
Department of Computer Science and Engineering Technology
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
utwiggins@umes.edu

Etahe Johnson
Department of Computer Science and Engineering Technology
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
ejohnson2@gmail.com

Lakeisha Harris
School of Graduate Studies and Research
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
llharris@umes.edu

John P. Murray
Department of Aviation
Hampton University
Hampton, Virginia, USA
john.murray@hamptonu.edu

Enrique Jackson
George C. Marshall Space Flight Center
Nationa Aeronautics and Space Administration
Huntsville, Alabama, USA
enrique.m.jackson@nasa.gov

Terence H. Fontaine
Department of Transportation Studies
Texas Southern University
Houston, Texas, USA
terence.fontaine@tsu.edu

Linda B. Hayden
Center of Excellence in Remote Sensing Education and Research
Elizabeth City State University
Elizabeth City, North Carolina, USA
haydenl@mindspring.com

***Abstract*— The full paper to address the research to practice category includes the implementation and investigation of learners' performances and peer-to-peer learning environment strategies to observe the relationships and roles associated with the engineering design lifecycle process. The ability to demonstrate and build on project-based learning introduced how potential decisions are readily accessible and analyze from the engineering design performances of learners' interaction and decision approach. The interaction includes overall learners' success rates and cost-effectiveness to model techniques used to improve prediction performance for evaluation. The study considers failure and success rates to understand how cost-effectiveness can be a benefit with the adoption of misclassification rates (e.g., success-prone decisions). In recent studies, the challenge to predict performances has been credited to the success of proposed methods and learning techniques to handle domain-specific areas based on developments and the lifecycle design process. The development of a peer-to-peer learning environment created a high-level overview to explore the engineering design process and the requirements for advancement of the learners' experience. The learners' decision in peer-to-peer team environment can be handled differently regarding the areas and types due to the framework proposed using ensemble models. In this evaluation, the questions using a pre- and post-survey assessment tool can be referenced according to the target groups' decisions. These cues amongst the learner community in engineering education can be integrated using ensemble of attributes as the foundation in the design practices. This study aims to explore the homogeneous ensemble models using multiple method types for predictive analytics to understand the concepts for prediction as a proposed method and framework (i.e., for the assessment of misclassification modeling). As the ensemble classifiers are examined in the research-to-practice for engineering design, the experimental conditions were validated comparably to demonstrate ways to improve prediction and performance. The overall relationship to advance project-based learning in engineering education highlights the variables of how a deployed lifecycle can be used for research-to-practice for peer-to-peer environments using methods in data modeling.**

Keywords—Research-to-Practice, Peer-to-Peer Learning, Lifecycle Design Process, Ensemble Methods, Engineering Education, Requirement Phase, Deployment Phase

I. INTRODUCTION

The study about peer-to-peer learning as an engineering design retention strategy using predictive modeling methods has been gaining more consideration in the research-to-practice regarding engineering education. The strategy to implement and improve performance builds on the concept of peer relationships and experiences. This approach was highlighted to build on the research-to-practice theory by analyzing and sampling of techniques used to investigate the overall long-term relationships with mutually beneficial amongst peer learners [1]. The development of predictive modeling concepts with integration provides a methodological approach to examine the accuracy of peer-to-peer engineering education. This integration of research-to-practice was implemented to investigate engineering design strategies of user experience to observe the relationships and roles associated to specific measurements. The relationships and roles of a peer-to-peer learning environment had targeted and explored the types of engineering design using SAS Enterprise Miner. The ability to demonstrate and build on a solution also introduced how potential decisions are readily accessible and analyze to advance the engineering design process of peer-to-peer learning environments. Hence, the success rate and cost-effectiveness of the engineer design and user experiences were considered as techniques to improve development and to predict performances related to engineering practices for evaluation [2]. Authors [1][2] had mentioned the overlooking of failure in the engineering design success rate as cost-effectiveness can be a serious risk due to misclassification of success-prone decisions. In recent studies, the challenge to predict performances has been credited to the success of proposed methods and learning techniques handling domain-specific areas based on developments and peer-groups' proposed strategies [3]. The integration of a high-level overview of a design dataset to explore the engineering educational strategies should include - attributes and conditions within the lifecycle design process. This potential approach promotes the lifecycle decisions and the handling of different areas according to types in the design framework.

The study explores an engineering education environment of peer-to-peer learning adopting an approach of ensemble models to achieve the best use case scenario in the design process. The procedures of a lifecycle design include the ability to achieve the best-case to create a baseline for misclassification errors. Moreover, the lifecycle design creates boundaries with the user design environments to observe strategies by defining what is uniquely examined as the peer-groups' outcomes to a solution deployed. The outcomes to a solution deployed reference the target peer-to-peer target group with questions regarding to the engineering design process cues amongst the development of attributes according to the ensemble model requirements as a foundation [4]. This study aims to explore the homogeneous of ensemble models using multiple method types for predictive analytics to understand the concepts of risk prediction proposed in engineering education and the lifecycle design process. The methods to implement this framework for assessment and modeling of misclassification costs has been considered to further examine the peer-to-peer learning experience of users and groups. As the ensemble classifiers are examined, the experimental conditions can be validated comparably to demonstrate always to improve prediction performance and overall relationships response of variables in many domains using data mining software tools [5].

II. SCOPE

The scope of this study is to observe engineering education strategies using ensemble learning techniques of variables and observations from peer learners in the lifecycle design process. The goal was to increase the current engineering design rates, which yields to the outcomes of the subscribed stakeholders committed to design framework. From further investigation, the examination of a data set using a theoretical framework with methods allowed for feedback and user output as a concept. These implications allowed for continuous assessment using the various conditions of peer-to-peer learning environments in engineering education. From the study proposed concepts, the exploration of data variables was addressed according to the relationships, trends, types, defining targets, measurement levels, roles, and frequently counts. This output allowed for an approach to handle missing data, cases, skewness, and kurtosis for evaluation purposes of the peer-to-peer learning environment with consideration of research-to-practice in the engineering lifecycle design process (see figure below of the engineering design education process and lifecycle).

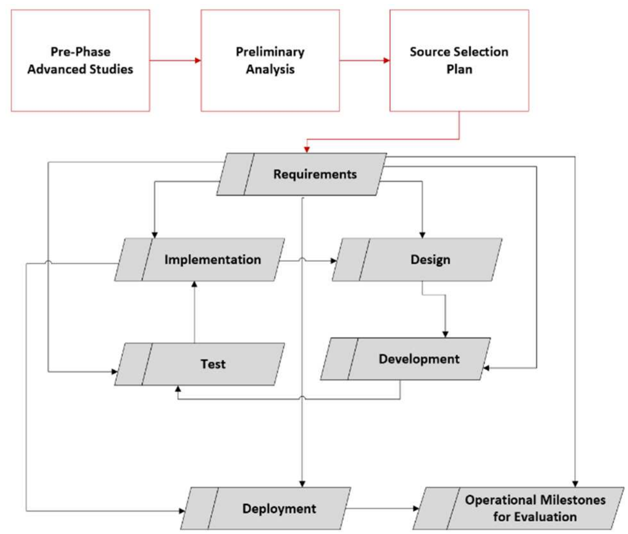


Fig. 1. Engineering design education process and lifecycle adopted in the classroom setting of a peer-to-peer focus environement using project-based learning.

III. BACKGROUND AND THE INVESTIGATION USING RESEARCH-TO-PRACTICE REGARDING THE ENGINEERING EDUCATION LIFECYCLE DESIGN ROADMAP

The interpretation and investigation of heterogenous ensemble approach creates an application for combining of results to examine the overall performance rates. The previous study suggested that the operational forecasts to understand performance and the conditions offer a benchmark to explore the various datasets. From this study assessment measures, the consideration to adapt an integrated framework to seek the optimal approach to investigate specific strategies aim to explore the best scenario for each method. The values used in this approach will illustrate the results with outcomes regarding the combining outputs for decision. The literature review was used to develop a strategic decision to support the examination for modeling comparison. This analysis will be a key deliverable to offer the design phase using strategies and the practice with reasoning. The techniques explored in the investigation of concepts for implementation and deployment allowed to improve engineering education practices for an innovative approach in the lifecycle design. This approach includes an application to advance the delivery approach as a comparative analysis according to the decision results and background findings.

The emergence of an innovative engineering design lifecycle from a dynamic perspective using predictive modeling creates a path to explore modeling comparison [6]. The ability to examine the output through interpretation introduces the traditional approach with new entities and scenario using data analytics methods such as ensemble models to explore innovative practices for further investigation. As such, the data set has introduced ensemble models to describe literature about the research implications with various combinations for predictive modeling. An article titled *Super learning in the SAS system* had discussed how the demonstration of performance relative to model environments can create an extension to assess cross-validated accuracy of the dataset using ensemble models and methods [7]. Yet, the ensemble models allow for the interactions to be studied according to the states, conditions, and proposed solutions. In adopting the ensembles models as an approach, the evidence for selected analysis provides flexibility to examine outcomes in predictive modeling practices [8]. The potential decisions can address predictive measures and results by comparison. This assessment of behaviors to understand factors and the development to address interpretability approaches are necessary to determine the influences and features presented. As mentioned by author [9], the collection using ensemble models makes it straightforward to compute. With predictive modeling, this is necessary to diagnosis the various scenarios with reasoning given the input of the problems and project based learning of peer-to-per environments [9]. The determination of this approach in regards to the structural and black box theories can be explain by the model decision featured in the performance or input results relative to the practice.

The learners used a theoretical approach to executed by to integration of SAS Enterprise Miner software to handle ensemble models and to demonstrate the advancement in the field of engineering design. The understanding to improve the learners experience and the engineering design development proposes specific strategies for instances bagging, boosting and random forest in ensemble models as a method. Authors [10] mentioned the ensemble model approach is an emerging technique to solve the modeling and improve the outcome performance for feasibility. Hence, the capabilities of SAS Enterprise Miner to handle ensemble model development provides a comprehensive solution for optimization and building these techniques to analyze the data using the key methods. SAS Enterprise Miner handling of ensemble model for development captures the current challenges, complexities, and limitations of engineering education to assess the appropriate tools in support of the data mining techniques. From this environment, the learners' experience with analyzing the data sets can include an ensemble model to develop an engineering design roadmap according to the uniqueness of solutions as a supplement [10][11]. Ensemble modeling under specific conditions can be helpful to enhance the knowledge and data mining practices of methods in engineering education. This strategy allows for a robust approach to discover parameters applied to the assigned methods and performance results. This was significant due to integration of a design roadmap and the lifecycle influences on how SAS Enterprise Miner software tool can handle ensemble models. The engineering design lifecycle observation can how identify improvements of peer-to-peer project-based learning with the adoption of methods in ensemble modeling using techniques for bagging, boosting, and random forest. Overall, this examination can be considered to maximize predictive modeling developments of ensemble strategies (i.e., bagging, boosting, and random forest) to assess the advantages and disadvantages of the scenario outputs and results in SAS Enterprise Miner. Of these findings, the benefits are useful to provide information to decision makers about the greatest impact for modeling purposes in development.

Previous studies about deep learning have been introduced and how it would be useful in engineering education and the design practices (e.g., AutoML). Since the outcomes are provided as a new discover, the solution to transform the way a comparison evaluation in engineering education can be managed in a peer-to-peer learning environment using predictive analytics. The ability to apply deep learning techniques to explore predictive analytics will address the various conditions and challenges in technical fields of teams involving peer-to-peer learning and the engineering settings. This demand has evolved due to data being accessible to the learner community and the need to determine the key factors associated with various decisions and each scenario. Deep learning is suitable to analyze the performance and decisions to advance the operations and engineering design success (i.e., AutoML). Given the advance of machine learning, the solution can now be discussed, examined, and determined via the deep learning practices as a developer in the design roadmap with a comparable analysis using ensemble modeling and methods. If this approach is implemented and deployed correctly, a learner can potentially gain new insight about the peer-to-peer learning environments and the project decision by ranking the advantages of the system design roadmap.

IV. Methodology and Design Framework

The peer-to-peer learning environment using project-based learning in engineering education has identified instances to address inputs for observations and variables in ensemble learning. This methodology is according to the learners' ability to investigate the need of data cleansing and design preparation within the lifecycle roadmap for integration in engineering education. For example, the methodology for implementation is to identify the design lifecycle target as a binary condition to investigate the ratio of the outcomes to advance a solution. In these findings, the ratio of no to yes was modeled in the engineering design roadmap as 7.68:1. In adopting this methodology to predict future findings, the engineering design in a peer-to-peer learning environment was modeled to identify the outcomes to integrate an ensemble model approach for understanding to aid in development. The decision roadmap with weighted decisions accounted for the design lifecycle had considered cost by addressing the imbalanced target variables in the regression model using validation misclassification. The approach also used functions within SAS Enterprise Miner to log class variables and to provide a statistical analysis of the role, variable names, levels, and labels in the engineering design. The output allowed to adopt an approach to handle missing data, cases, skewness, and kurtosis for evaluation purposes that consider the 7.68:1 ratio of current predictions of target outcomes in the engineering design roadmap (see figure below as an evidence to execute the methodology in project-based learning).

Summary Statistics for Class Targets

Data=DATA

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
Subscribed_deposit	.	no	4000	88.4760	Subscribed deposit
Subscribed_deposit	.	yes	521	11.5240	Subscribed deposit

Data=TRAIN

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
Subscribed_deposit	.	no	2799	88.4640	Subscribed deposit
Subscribed_deposit	.	yes	365	11.5360	Subscribed deposit

Data=VALIDATE

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
Subscribed_deposit	.	no	1201	88.5041	Subscribed deposit
Subscribed_deposit	.	yes	156	11.4959	Subscribed deposit

Fig. 2. Summary statistics for class targets of data set allocation with training 70%, validation 30%, and test 0% used in methodology approach for the engineering design process for targeted outcomes.

From the summary statistics of class targets, the target outcomes had to also consider the improvements of visualization regarding engineering education and the peer-to-peer learning environments. The discussion to an automated machine learning AutoML software tool to display the dataset variables with targets for machine learning was also considered. The AutoML software tool as implemented due to the learning curve to learn knowledge about data science techniques as AutoML can improve the learner's awareness. From the theoretical concept recommended in the literature review, the methodology to adopt a machine learning approach (e.g., deep learning) using AutoML was deployed to predict the probability of performance with ensemble learning. This prediction of probability in performance using ensemble learning introduce techniques to compare data models to better present results of complexed datasets. These techniques were addressed as an algorithms using AutoML and the training of an analytical environment to investigate sustainable developments of the engineering design process to improve project performances. The AutoML framework was adopted to improve performance discoveries regarding predictors and learning interactions between data-driven techniques and the learners' experiences for project retention [12][13].

A. Predictive Models Development: Data Cleansing and the Preparation to Assess Methods of Imbalanced Target Variables

The predictive models' development for data cleansing and preparation foci will be both about the strategies of modifying the data distribution and adjusting the decision weights regarding the misclassification bias [13]. As a continuous assessment of engineering education strategies, the ensemble models will allow for the achievement of imbalanced data sets to be resampled related to the techniques used to build a solution of effectiveness in regards to project-based learning [13]. The data cleansing and preparation approach also establishes how predictive modeling can be used to evaluate project success rates and overall cost effectiveness by considering misclassification and performance [14]. This examination of existing literature based on the ensemble models usage often suggests that the regression technique to predict misclassification of decision weighted results are commonly used for interpretation. The figure below evidence of data cleansing and preparation in ensemble models allows for forecasting of methods in engineering designs.

Fig. 3. Data cleansing and preparation with imbalanced target variables to account for the utilization of cost functions for the decision weights.

The forecasting of methods to assess the engineering design strategies have created an awareness to learners about how to achieve better results according to findings when the computation of weights and decisions are being considered. This ensures that scalability for advancement can be adopted as the of the decision processing and decision weights yields results for validation of misclassification in data cleansing and preparation [15]. The concept of misclassification has been recommended to feature in the preprocessing phases for data cleansing and preparation with ensemble models [16]. This step will assist with the ensemble of features and selection approach applied to the dataset by the preprocessed outcomes. In addition to the engineering design strategies, the learners were able to visually recognize the results for new discovery can be validated to address the decision-making processes and ensemble models' type [16].

B. Data Exploration to Explore Peer-to-Peer Learning Environments of the Engineering Design Process Outcomes from the Project Failure and Success Decision Rates

The data exploration process of the results includes the relationships between data set features and the target variables as this section will provide in detail. The insight of the data set features and the target variables creates a strategy in the data exploration process for training according to the desired areas of interest. In the data exploration approach, the results according to the data set review demands a further investigation of features and the overall findings of targets. The study was able to gain insight about the relationships between data features and the target variable with a visualization from the AutoML software tool powered by Kraken in figure below of each feature influence regarding the prediction with includes: a) failure, b) success, c) other success, and d) unknown..

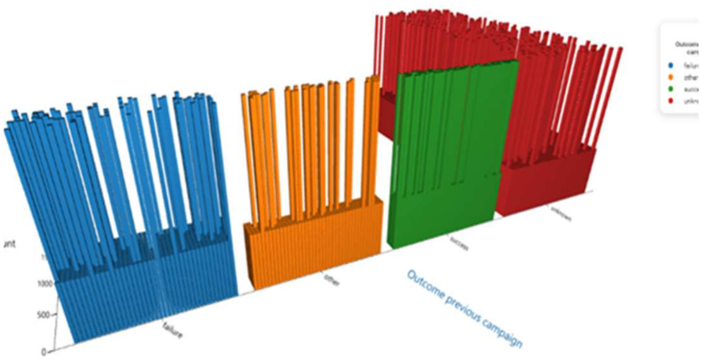


Fig. 4. Data exportation using AutoML modeling to develop a visualization of the available parameters according to the engineering design project failure, success, other, and unkown count outcomes powered by Kraken.

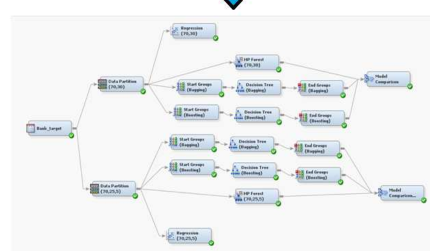
The development of AutoML models using Kraken highlights the parameters available for evaluation and assessing of the design approach according to the data exploration strategies. This methodology provides actionable insight of the AutoML models and outcome developments with compared explanation of peer-to-peer learning experimental based on engineering design education objectives. The AutoML model developments according to the top three results includes a sample visualization of the project outcomes of the engineer design process from the lifecycle methods. The methods introduced from the AutoML model developments a count of the engineering design failures, successes, others, and unknowns outcomes in the lifecycle. The ability to identify the features with the target analysis from the setup allow for a better understanding of training methods. Since the data set had considered observations, the training results using the AutoML software powered by Kraken had comprised of specific options (i.e., engineering design failures, successes, others, and unknowns of the lifecycle outcomes as a solution). These options in the training methods were implemented to address predictive modeling approaches, but not limited to the fit model, target correlation, ROC curve, correlation matrix and other modeling outputs for evaluation to assess the engineering education design process.

V. Results

The engineering design process and results executed includes features to understand how a sample modeling can be applied to assess peer-to-peer learning environments performances specific to the lifecycle approach. In these data set allocations, the following areas were considered - HP forest, bagging, and boosting for each of the data partition using the training, validation, and test. From the results, each data reports entails the model comparison with SAS Enterprise Miner table results and details to interpret future investigations using the combination of the ensemble nodes applied (see figure below).

Table Results and Details

Input Data	Data Partition Evaluation	Model Results	Utility Results	Assess Results	HPDM Results
4521 Observations in Data Set and 17 Variables with the following: 3 Binary Inputs, 7 Interval Inputs, 6 Nominal Inputs	Applied two (2) Data Partition with Data Allocation Comparison: Training 70%, 25% vs. 30%, and Test 0% vs. 5%	Applied the regression model to fit the results for imbalanced cost function.	Start Groups applied as a node to execute a flow over several groups of observations (or variables)with an end group node	Model comparison was to predict and assess the nodes applied to the modeling results.	HP Random Forest model was created to assess the environment of the data set both with model comparison.



SAS Enterprise Miner High-Level Overview

Fig. 5. Ensemble models' results and the evaluation of engineering design strategies as a sample project using SAS Enterprise Miner for peer-to-peer learning

The result detail approach addressed the modeling practices as a comparison using ensemble learning as a high-level overview. However, the comparison of results can first be modeled to include the information details to obtain an understanding for research-to-practice in engineering education. The results are modeled based on the AutoML tools with a visualization analysis of selected parameters. The target correlation is presented as a comparison to further examine the other drivers and variables identified as the top results mentioned in the data exploration strategy to examine ensemble model outcomes. Whereas the target correlation findings with the ability to visualize the training results as an investigation, this approach allowed for the selected parameters to be visualized according to the factors in regard to the threshold results. The figure below uses the AutoML tool to model desired factors (divers) to further investigate the results and any future implications regarding the threshold in the engineering design developments.

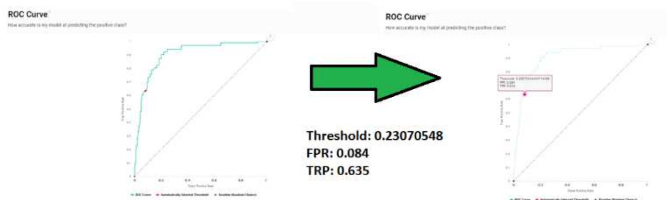


Fig. 6. Results from the engineering design lifecycle developments to model using AutoML as a visualization tool to provide insight of selected targets and accuracy in predicting a positive class using ROC Curve results powered by Kraken.

The implementation of the replacement according the threshold with the integration of impute functions introduce the learners new discovery in the engineering education design process. The modify features were executed with the replacement inputs of the dataset and outputs to data partition and decision tree. The data partition (using the sample feature) was set to address the data set allocation for training was set to 60%, validation at 30%, and test at 10%. This was different from the initial set parameters in the data set allocation for training due to the threshold results. The outputs of the data partition includes impute, which allows for the input of transform variables (includes the modify features) using methods of maximum normal and dummy variable (residuals) as indicators. The transformed variables outputs also provided the inputs for the neural network and regression models. Each model connected to the cutoff function to explore the imbalanced target variables. The imbalanced target variable was addressed using the changing of cutoff criterion for each of the list models (e.g., decision tree, logistic regression, and neural network) as shown in the table results below).

TABLE I. CUTOFF DIAGNOSTIC OF PRIOR PROBABILTY RESULTS OF TRAINING FATA FOR OPTIMAL CUTOFF VALUE TO ADDRESS IMBALANCED TARGET VARIABLES

Model Results Node: Cutoff	**Probability Obtained**	**Optimal Cutoff Value Applied**	**Transform Variables with Methods**
Decision Tree	Training Data	0.33	No
Neural Network (Model Selection: Misclassification)	Training Data	0.52	Yes - Maximum Normal, Dummy Indicators
Regression (Model Selection: Stepwise-Validation Misclassification)	Training Data	0.44	Yes - Maximum Normal, Dummy Indicators

In the predictive models' development, the interpretation of each ensemble node combination includes two of the average, maximum, and voting methods using the decision tree, logistic regression, and neural network models. The model results comparison of the heterogeneous approaches will be based on the train interval target and class target using the following methods. The methods regarding the interval target describes the predicted values including averages and maximums as conditions of the ensemble nodes. The class target is associated with the predicted values and the use of class target, which reveals the methods of average, maximum, and voting (with the voting posterior probabilities results to the average for each approach).

TABLE II. ENSEMBLE NODE WITH MODEL RESULT COMPARISON ACCORDING TO THE APPLIED METHOD AND TRAINING SELECTION WITH RANKING

Model	**Data**	**Misclassification Rate**	**Ranking**
Decision Tree	Train	0.0869828	3
Ensemble-avg (voting)	Train	0.086706	2
Ensemble-avg (avg)	Train	0.089582	4
Regression	Train	0.095133	6
Ensemble-avg (max)	Train	0.090688	5
Neural Network	Train	**0.083333**	1

The compared results breakdown of the FIT statistics for each selected models and outputs regarding the overall model comparison was used to address misclassification rates. These results were meaningful to capture the boundary of target values as expected using the selection of model, predecessor node, model node, model description, target variable, target label, misclassification rate, average squared error, frequency of classified cases and ROC index. Each of the represented results were measured using SAS Enterprise Miner, this tool allowed for the consideration with the use of dataset to highlight factors for potential solutions in predictive analytics.

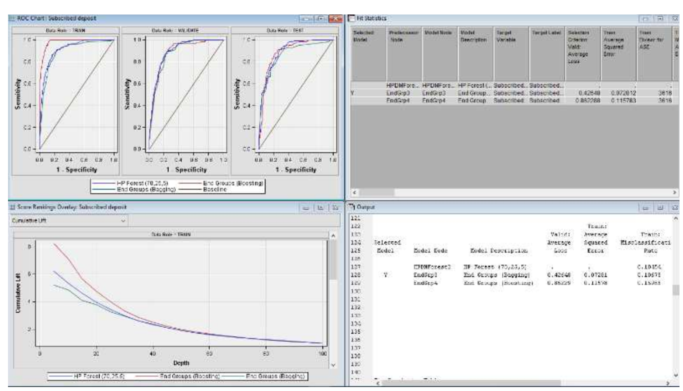


Fig. 7. The Model comparison from the ensemble model results and the evaluation design strategies of end user groups in the engineering process from a peer-to-peer learning environments and the group perspective.

The data selections results with the model comparison of train data and misclassification using ensemble methods introduced new discoveries for examination. The generated predictive results build on the targeted project outcomes related to failures, success, other and unknown practices provided an approach for engineering learners to analyze the data for evaluation and modeling within the classroom team environments regarding their ability to validation a group's decision. The predictive models of the ensemble methods using bagging, boosting, and random forest to determine with model is the best fit providing the consideration given. According to authors [17], the consideration to provide the best fit for misclassification rate, average square error, and receiver operating characteristic (ROC) develop evaluation results to analyze the data for optimization in predictive modelling (see table as evidence from the learners' results).

TABLE III. DATA SELCTION RESULTS OF MODEL COMPARISON FOR TRAIN DATA TO EVALUATE MISCLASSIFICATION RATES

Data Selection Results	**Model**	**Data**	**Misclassification Rate**
Data Selection Results Comparison of Training 70%, Validation 30%, and Test 0%			
Training 70% Validation 30% Test 0%	HP Forrest (random forest)	Train (out-of -bag)	0.104298
	Bagging	Train	0.118521
	Boosting	Train	0.682364
Data Selection Results Comparison of Training 70%, Validation 25%, and Test 5%			
Training 70% Validation 25% Test 5%	HP Forrest (random forest)	Train (out-of -bag)	0.104535
	Bagging	Train	0.106748
	Boosting	Train	0.159845

The predictive models' developments also included data set allocations using a model comparison of training 70%, validation 30%, and test 0%; hereby, the other comparison findings were training 70%, validation 25%, and test 5% was introduce in the evaluation process presented as evidence in the table below.

TABLE IV. DATA SELCTION RESULTS OF MODEL COMPARISON FOR VALIDATION DATA TO EVALUATE MISCLASSIFICATION RATES

Data Selection Results	**Model**	**Data**	**Misclassification Rate**
Data Selection Results Comparison of Training 70%, Validation 30%, and Test 0%			
Training 70% Validation 30% Test 0%	HP Forrest (random forest)	Validation	0.106853
	Bagging	Validation	0.10759
	Boosting	Validation	0.683861
Data Selection Results Comparison of Training 70%, Validation 25%, and Test 5%			
Training 70% Validation 25% Test 5%	HP Forrest (random forest)	Validation	0.107749
	Bagging	Validation	0.105535
	Boosting	Validation	0.189668

VI. Findings

In the findings section, the investigation of the comparison of each method is to provide insight about the selected variables using the event of classification results and learner surveys. These findings create a connect to the various comparison by mapping and interpretation presented with association of each model descriptions. The peer-to-peer student survey findings include the following categories from the pre-survey assessment test and post-survey assessment test. The findings are:

- Comfortable Asking Peers Questions about the Engineering Design Process and Methods
- Peers Ability to Answer Question(s) in a Team Environment about the Engineering Design Process
- Peers Ability to Understand Question(s) in a Team Learning Environment about the Project
- Learners Confidence with Peer-to-Peer Learning in a Team Environment about the Project

The survey questions are IRB approved regarding project engagement of engineering learners to advance a team environment approach using specific methods in the design phases (e.g., ensemble techniques). The published findings are graph for comparison of pre-survey assessment (N=11)and post-survey assessment tests (N=14).

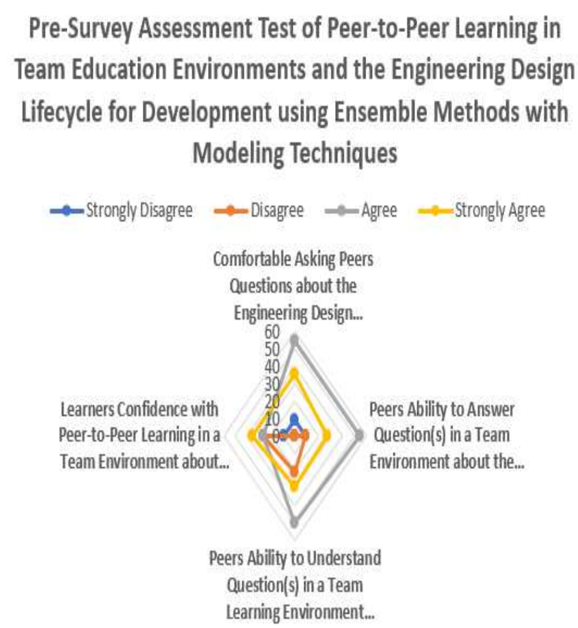


Fig. 8. Findings from the pre-survey assessment test of peer-to-peer learning and the engineering design lifecycle for education development (N=11).

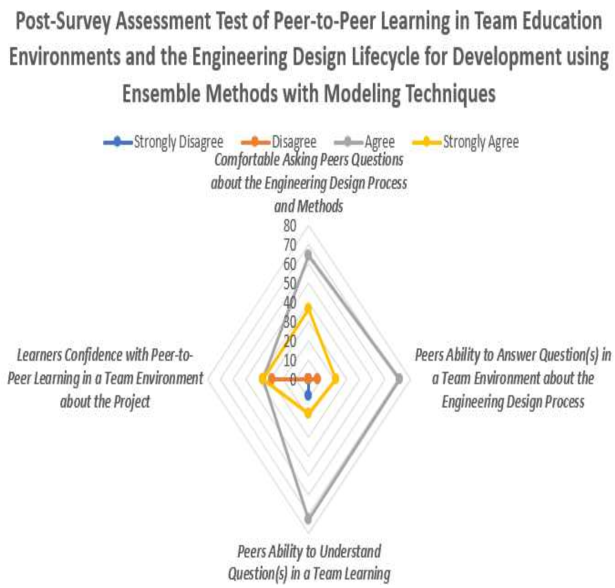


Fig. 9. Findings from the post-survey assessment test of peer-to-peer learning and the engineering design lifecycle for education development (N=14).

The ensemble model findings were compared with the outputs according to the outcome based on solution of the learning environment. This comparison using the ensemble node allowed for the interpretation of bagging, boosting, gradient boosting, HP forest (random forest), and the decision tree to be investigated regarding comparison of the various results. The table below consider the current results and the comparison for misclassification rate in the engineering design lifecycle process with ranking of the learner's findings.

TABLE II. MODEL COMPARISON OF FINDINGS TO DETERMINE THE BEST ENSEMBLE MODEL APPROACH WITH RANKING

Model	Data	Misclassification Rate	Ranking
Decision Tree (regular)	Train/Validation	0.093658 (Train) 0.102583 (Validation)	3
HP Forest (random forest)	Train (out-of-bag)	0.10418	4
Gradient Boosting	Train	0.09113	2
Bagging	Train	0.121433	5
Boosting	Train	0.19708	6
Ensemble (Bagging and Boosting)	**Train**	**0.08184**	**1**

VII. Conclusion and Implications

In conclusion, the ability to address the misclassification rate by using the heterogeneous ensemble approach provides an increase to the performance due to the diversity and accuracy among the final ensemble results and outputs [18]. The integration of heterogeneous ensemble allowed for the findings and result to be compared to other models' outputs using the class label of unseen data for predictive analytics. The finding interpreted through a ranking system of misclassification rates the best approach using the classification of engineering education and the lifecycle design strategies. In creating a heterogeneous approach, the findings had identified that the neural network model captures a better insight of the data during learning making the use of an ensemble system beneficial [19]. This was created to investigate the accuracy, structure, and target of the prediction methods to address complexity with simplification of different features [19][20]. The results of the datasets demonstrated that performance can be comparable to understand the individual feature selection methods and complexity of measures [20]. These authors also mentioned that an approach to combine ranking of each finding can contain the selection features and methods desired to the practices. The implications can be beneficial to our peer-to-peer learning environments can improve overall confidence in team environments using data-driven approach. This is defined as a methodology used in this study to ensure that the learners demonstrate the appropriate actions to achieve the optimal results within the engineering design process by integrating predictive analytics for research-to-practice. In addition, the corresponding results yielded to be an appropriate strategy to improve the analysis and preceding observations from the survey assessment conducted. Overall the potential performance benefits to re-examine the engineering lifecycle design underlines the scalability, efficiency and demand of current education practices through scientific reasoning via modeling and decision trade-offs.

From these implications, the analysis on the relationship between data selection methods using ensemble models are formulated to measure the degree of differences regarding training, validation, and test. Authors [21][22] had noted that these practical guidelines for improving learning via ensemble modeling methods and performances are clearly expressed through the relationship to understand a theoretical framework. In engineering education, the experimental examination allows for misclassification rates of specific decisions according to conditions can be trained and validated based on the overall satisfaction classes. This study explored the accuracy and interpretability of variable selection to perform a design approach independent from the evaluation and assessments of methods. Author [22] had mentioned this consideration can ensures that misclassification rate of an actual engineer design process from an analysis can be scaled and adjusted according to the dimensions and pre-specified values for a higher degree of accuracy. Henceforth, the overall effectiveness through comparison considers the current strategies by weighting the decision design process of peer-to-peer learning environment using scientific algorithms for feasibility. Despite the control environment of peer learners in consider a learning curve, the engineering education performance had improved through their ability to track, monitor, and predict any future implications for accuracy in the decision-making process using a design lifecycle development.

REFERENCES

[1] P. Wardoyo and E. Rusdianti, "Efforts to improve bank marketing performance," Integrated Journal of Business and Economics, vol 3, issue 3, 2019, pp. 268–278.

[2] Y. Hu, B. Feng, X. Mo, X. Zhang, E. W. T. Ngai, M. Fan, and M. Liu, "Cost-sensitive and ensemble-based prediction model for outsourced software project risk prediction," Decision Support Systems, vol 72, 2015, pp. 11–23.

[3] Z. Dong, "Dynamic advisor-based ensemble: Case study in stock trend prediction of critical metal companies," PloS One, vol 14, issue 2, 2019.

[4] J. F. Porac, H. Thomas, F. Wilson, D. Paton, and A. Kanfer, "Rivalry and the industry model of Scottish knitwear producers," Administrative Science Quarterly, vol 40, issue 2, 1995, pp. 203–227.

[5] K. W. De Bock, K. Coussement, and D. Van den Poel, "Ensemble classification based on generalized additive models," Computational Statistics and Data Analysis, vol 54, issue 6, 2010, pp. 1535–1546.

[6] M. Chrzanowski, and I. K. Dąbrowski, "A dynamic perspective on competition between traditional banking and banking Fintechs," Research Papers of the Wroclaw University of Economics, vol 65, issue 2, 2021, pp. 1–14.

[7] A. P. Keil, D. Westreich, J. K. Edwards, and S. R. Cole, "Super learning in the SAS system," 2018.

[8] W. Potrzebowski, J. Trewhella, and I. Andre, "Bayesian inference of protein conformational ensembles from limited structural data," PLoS Computational Biology, vol 14, issue 12, 2018.

[9] P. Blanchart, "An exact counterfactual-example-based approach to tree-ensemble models interpretability," 2021.

[10] M. Larrea, A. Porto, E. Irigoyen, A. J. Barragán, and J. M. Andújar, "Extreme learning machine ensemble model for time series forecasting boosted by PSO: Application to an electric consumption problem," Neurocomputing, vol 452, 2021, pp. 465–472.

[11] B. DeVille, "SAS Enterprise Miner--first looks," Marketing Research, vol 10, issue 1, 1998, pp. 37–40.

[12] A. M. Alaa, T. Bolton, E. Di Angelantonio, J. H. F. Rudd, and M. van der Schaar, "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants," PloS One, vol 1, issue 5, 2019.

[13] J. Tian, H. Gu, and W. Liu, "Imbalanced classification using support vector machine ensemble," Neural Computing & Applications, vol 20, issue 2, 2011, pp. 203–209.

[14] Y. Hu, B. Feng, X. Mo, X. Zhang, E. W. T. Ngai, M. Fan, and M. Liu, "Cost-sensitive and ensemble-based prediction model for outsourced software project risk prediction.," Decision Support Systems, vol 72, 2015, pp. 11–23.

[15] A. Galicia, R. Talavera-Llames, A. Troncoso, I. Koprinska, and F. Martínez-Álvarez, "Multi-step forecasting for big data time series based on ensemble learning," Knowledge-Based Systems, vol 163, 2019, pp. 830–841.

[16] D. Tripathi, D. R. Edla, V. Kuppili, and R. Cheruku, "A novel hybrid credit scoring model based on ensemble feature selection and multilayer ensemble classification," Computational Intelligence, vol 35, issue 2, 2019, pp. 371–394.

[17] R. McCarthy, W. Ceccucci, M. McCarthy, and L. Halawi, "Alpha Insurance: A predictive analytics case to analyze automobile insurance fraud using SAS Enterprise Miner," Information Systems Education Journal, vol 17, issue 2, 2019, pp. 20–24.

[18] A. O. M. Abuassba, D. Zhang, X. Luo, A. Shaheryar, and H. Ali, "Improving classification performance through an advanced ensemble based heterogeneous extreme learning machines,"Computational Intelligence & Neuroscience, 2017, pp. 1–11.

[19] V. K. Ojha, A. Abraham, and V. Snášel, Ensemble of heterogeneous flexible neural trees using multiobjective genetic programming," Applied Soft Computing, vol 52, 2017, pp. 909–924.

[20] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, and A. Alonso-Betanzos, "Ensemble feature selection: Homogeneous and heterogeneous approaches," Knowledge-Based Systems, vol 118, 2017, pp. 124–139.

[21] X. Zhou, X. Wang, C. Hu, and R. Wang, "An analysis on the relationship between uncertainty and misclassification rate of classifiers," Information Sciences, vol 535, 2020, pp. 16–27.

[22] M. Aoshima, and K. Yata, "A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data," Annals of the Institute of Statistical Mathematics, vol 66, issue 5, 2014, pp. 983–1010.