Journal of Engineering, Computing and Technology
Volume 1, Issue 1
Journal Homepage: https://www.isujournals.ph/index.php/ject
Publisher: Isabela State University, Echague, Isabela, Philippines

Prediction Model on the Relationship of Undergraduate Grades and Licensure Examination Performance of BS Agriculture and Biosystems Engineering

Kim Darell Rosal Salgado[1], Edwin Romeroso Arboleda[2], Mariella Janelle Dime Cabardo[3], Chenna Mae Peligrino Obejera[4], Jesusimo L. Dioses, Jr.[5]
Department of Computer and Electronics Engineering, Cavite State University, Indang, Cavite, Philippines[1,2,3,4]
College of Computing Studies, Information and Communication Technology, Isabela State University, Echague, Isabela, Philippines [5]
edwin.r.arboleda@cvsu.edu.ph

RESEARCH ARTICLE INFORMATION

Received: May 26, 2023
Reviewed: June 19, 2023
Accepted: June 05, 2024
Published: June 29, 2024

ABSTRACT

This study employed WEKA software and data mining techniques in order to identify the critical grade subject(s) needed for passing the Licensure Examination for the BS Agricultural and Biosystems Engineering. The study's proponents examined the ABE licensure examination and grades of 84 BSABE graduate students between September 2015 and October 2019 in order to ascertain if higher scores on undergraduate exams correlate with passing the licensure examination. Researchers also created a dataset covering all academic subjects within the BSABE Program, such as Engineering Mathematics, Science Subjects, Major Subjects, General Subjects, and Competency Appraisal subjects. Their results suggested that undergraduate students who scored highly on these subjects during their undergraduate studies stood a better chance of passing the ABE Licensure Examination.

Keywords: *attribute evaluator, classifier, data mining, prediction model, ranker attribute, WEKA*

Introduction

According to Parnell and Carley (2015), modern engineering poses many complex challenges to its practitioners, such as ethical and social responsibilities, sustainable development, and effective communication. As our economy is moving towards development, many other things are downsizing it (Bae et al., 2019). Engineers therefore require both technical and interpersonal communication skills in order to navigate this multidimensional environment of modern engineering effectively. As a way of meeting these challenges, many countries have implemented accreditation processes and licensure examinations in order to certify engineering professionals as being industry-ready (Tamayo et al., 2014). The certification serves as a regulatory mechanism requirement before any professional can practice their profession (Gonzales et al., 2019; Bruce et al., 2019), and provides a foundation of standardization for professional engineers, engineering educators, and educational researchers (Tamayo et al., 2014). It is governed by applicable laws that establish the minimum level of expertise in a range of subjects. (Gonzales et al., 2019). Additionally, this assessment serves as a measure for professional practice analysis and plans to help enhance higher education curricula across the country's higher institutions (Tamayo et al., 2014). The Professional Regulation Commission's certification of competence is meant to ensure that a graduate has mastered a body of information, demonstrated a minimum level of professional competence, and acquired the necessary abilities in a certain specialization (Gonzales et al., 2019; Banluta et al., 2013; Ferrer et al., 2016). One indicator of the quality of training and teaching provided by educational institutions is the performance in licensure examinations (Ladia et al., 2020). As a result, several HEIs examine their alumni's licensure performance in order to evaluate curricular offerings and devise intervention strategies (Chan-Rabanal et al., 2016; Maaliw, 2021). HEIs' major goals are to deliver high-quality education to their students and to raise the level of administrative decision-making. Discovering information from educational data to analyze the major factors that may impact the student's performance is one approach to reaching the maximum level of quality in the higher education system (Rustia et al., 2018). Higher education benefits from data mining, particularly in the teaching and learning process (Baepler & Murdoch, 2010). When data mining techniques are used on educational data, new knowledge or models emerge (Delavari et al., 2008). Educational data mining, or EDM, is well-positioned to make use of a vast quantity of data mining research and apply it to educational challenges in learning, cognition, and evaluation (Tarun, 2017).

The BS Agricultural and Biosystem Engineering (BSABE) program at Cavite State University—Don Severino Delas Alas Campus CvSU is one of the university's priority courses. This program requires board license examinations for its graduates to be named licensed agricultural and biosystem engineers, and in recent licensure examinations, this university has produced top scores in the Agricultural Engineering board examination. According to the Professional Regulation Commission, CvSU's overall performance in the recent 2021 licensure examination reached a 57.89% passing rate, or 8 out of 11 takers from the university, while the national passing rate of the licensure exam is 36.42% or 507 out of 1,392 takers. In the September 2021 exams, CvSU came in eighth place. College academic performance has been shown to predict success on licensure exams, with higher grades translating into greater chances of passing them (Landrum & Harrold, 2015; Alkharusi & Al-Hinai, 2018; Sánchez-Martín et al., 2019).

Furthermore, in order to have a huge possibility of passing a licensure exam, some factors might be considered such as having study habits, preparation strategies, and personal characteristics that may influence exam outcomes. Sverdlik and Hall (2016) emphasized the role of consistent study habits and regular review in achieving better exam results. Pau et al. (2017) highlighted the importance of self-regulated learning, such as setting goals, monitoring progress, and employing effective study techniques in improving examination outcomes. A study by Zeidner and Klug (2013) revealed that students who utilized deep learning strategies (e.g., relating concepts, seeking understanding) rather than surface learning strategies (e.g., memorization) demonstrated higher exam scores.

Additionally, Rienecker and Stray Jorgensen (2019) emphasized the value of employing a variety of preparation strategies, including practice exams, self-testing, and concept mapping to enhance exam outcomes. Research by Richardson et al. (2013) indicated that personal characteristics, such as motivation, persistence, and self-efficacy, were associated with better exam performance. The purpose of this research was to determine if there is any significant relationship between the data grades of all engineering mathematics subjects, science subjects, general academic subjects, and specialized subjects of BSABE to the board examination performances.

As defined by the experts, machine learning (ML) is the study that gives computers the ability to learn without explicit programming. It makes the machines carry out certain tasks smartly through learning on their own. Moreover, it is known for its capability of dealing with 'big data', as big data means bigger accuracy as there is more to learn during the training of ML. The bigger the data, the higher the knowledge the machine learning can gain from training, and the higher the accuracy it can produce (Geollegue et al., 2022). Data gathering, morphological characteristic extrication, as well as their illustration, classifier/algorithm option and knowledge, and classifier testing are the utmost vital procedures (Garcia et al., 2020). Data mining is the process of extracting significant previously unknown, and possibly valuable information from large amounts of data (Aldowah et al., 2019). It is a technique of extracting a large amount of information from a sample of information. Previously, data analysts were in charge of this duty, but computers have changed that, and it is now more efficient than the statistical method.

On the other hand, data mining is a technique for detecting patterns and trends in vast amounts of data. Data mining activities include description, prediction, estimation, classification, grouping, and association. It is used by businesses to transform raw data into useful information (Bachhal et al., 2021). Moreover, this concept provides a scientific inquiry into the development of tools for uncovering novel sorts of data in educational contexts, intending to apply these approaches to better understand students and their learning environments (Su & Lai, 2021). The fundamentals of data mining are beneficial in higher education, particularly in the process of learning different subjects and specialties as well as determining and forecasting students' future academic achievement in school (Bachhal et al., 2021). Classification and prediction are fundamental duties in the context of data mining techniques used to analyze and extract insights from data. These tasks involve the application of various algorithms and methods to classify data into predetermined categories or to predict future outcomes based on discovered patterns and relationships in the data.

Additionally, classification algorithms are frequently used in data mining to designate instances to predetermined classes or labels based on their characteristics or features. This research made use of the IBk classifier. It is a classification algorithm variant of the k-NN algorithm that labels new instances based on the majority class of their k nearest neighbors in the feature space. KNN or K nearest neighbor is a method used for classification and regression (Rabe et al., 2019). It is one of the most fundamental and simple classification methods. When there is little or no prior knowledge about the distribution of the data, the KNN method should be one of the first choices for classification (Manalo et al., 2019). It assumes all instances are points in n-dimensional space. It is a versatile and adaptable classification technique, but it requires computational resources and cautious parameter selection.

Also, the primary benefit of the IBk classifier is its simplicity and adaptability to various data types and decision boundaries. Since it uses actual instances from the training data, it can manage intricate relationships and makes no assumptions about the distribution of the underlying data. Moreover, it supports incremental learning, as new instances can be added to the existing model without extensive retraining.

Prediction, on the other hand, involves making predictions about future events or outcomes using historical data. Frequently, regression algorithms are utilized in prediction models to establish relationships between input variables and the predicted outcome. In the context of prediction algorithms or models, a "ranker" is a form of algorithm or model designed particularly to rank instances or items based on their relevance or significance. Unlike conventional prediction models, which aim to explicitly estimate a particular outcome or target variable, rankers aim to determine the relative ordering or ranking of instances based on their features or attributes. They help prioritize results or recommendations according to their relevance to a given query or user preferences. In this investigation, a ranker prediction model is used to identify the most crucial aspect of BSABE. According to the findings of the previous study, this element has a significant direct association with licensure performance (Banluta et al., 2013; Maaliw, 2021; Tarun, 2017; Antonio et al., 2016; Dagdag, 2018; Laguador & Dizon, 2013). The examiner's cognitive capacities are nearly taken into account in numerous research studies in determining the subjects that affect the performance of the various disciplines. The evaluations of Electronics and Communication Engineering and Mechanical Engineering board examinees from the previous study, who either passed or failed the PRC examination, were obtained from the Registrar's Office or the Engineering Department. These evaluations included the computed weighted average for all subjects (Laguador & Dizon, 2013; Dagdag, 2018), as well as the General Weighted Average (GWA) in high school and college, and college admission scores. These cognitive abilities are associated with success on the Licensure Examination for Teachers. conducted by Antonio et al. (2016).

Lastly, the academic parameters investigated were high school General Point Average (GPA), college admission test, college GPA, general education courses GPA, and major courses GPA conducted in agriculture graduates of the Bulacan Agricultural State College conducted by Nicolas et al. (2020). The proponents of this study determined if there is a significant relationship between getting high scores per subject and passing the board examination using educational data mining. Also, they predicted the critical subjects that could affect the academic performance of the graduates in the Agricultural and Biosystem Engineer Licensure Examination and found the best classification in terms of accuracy using Waikato Environment for Knowledge Analysis (WEKA) software. Morever, this research aimed to provide the students with insights about the critical subjects in passing the licensure examination. Also, the faculty of the said university benefits from this study to improve the educational status within the specific course program, and school administrators and faculty members of the college department to implement regulations, thereby improving the performance of the students in the licensure examination. Lastly, this study would assist future researchers in further developing prediction models for academic performance in specific college courses.

The Comprehensive Theoretical Basis

Conceptual frameworks provide researchers with a visual representation of their understanding of how variables in a study relate to one another and expected outcomes in terms of the expected outcomes of research project (Swaen, 2019). Here it serves as an illustration of how their research project proceeded and which factors they considered while carrying it out.

The input for this project consisted of all subject grades for BSABE graduates compiled by the CvSU database and verified through the Professional Regulatory Commission database, along with board exam passers verified through the Professional Regulatory Commission database.

After preprocessing, cleaning, and sorting had occurred, the data were loaded into WEKA software, which offers machine learning algorithms specifically geared to data mining tasks. WEKA's widely popular classification algorithms such as decision trees neural networks support vector machines identify features that influence BSABE graduates performance on board exams. This framework gives readers a clear picture of research methodology and data analysis methods in order to achieve the project goals. By organizing ideas into conceptual frames, researchers can organize and identify key factors of study design, and account for all necessary variables during analysis, thereby helping readers comprehend how the study unfolded and executed.

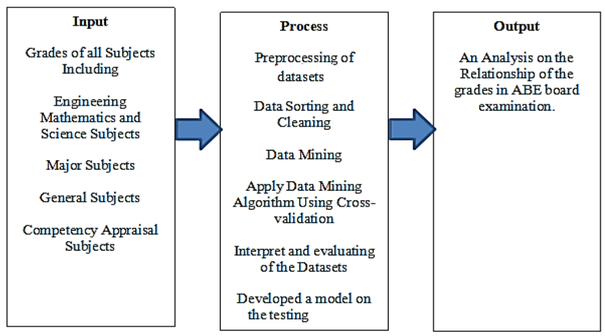


Figure 1. *Conceptual Framework of the Study*

Methods

Research Design

To investigate the grades of students enrolled in the Bachelor of Science in Agricultural and Biosystems Engineering program, a quantitative, quasi-experimental research approach was chosen. The goal of a quasi-experimental design is to identify a cause-and-effect relationship between an independent and a dependent variable. Quantitative research methods involve the process of acquiring numerical data and analyzing it using mathematical approaches, particularly statistics, to understand a problem or phenomenon. By studying the numerical grades of all participants enrolled in the years 2015–2019, the design was chosen as a means of determining the important relationship between high subjects scores in passing the licensure examination.

Dataset

The actual dataset was comprised of 84 first-time board exam takers at Cavite State University. Of the first-time takers of the board exam, 58 passed on their first attempt, while 26 failed. The ABE licensure examination was conducted from September 2015 to October 2019. The study excluded graduates who had not yet taken the licensing examination and those who had taken the examination more than once. There were also 75 attributes for each person, which corresponded to the 74 subjects of the BS in Agricultural and Biosystems Engineering. There were engineering mathematics subjects as well as general science subjects. Also, there were general academic subjects and specialized subjects (for BS ABE) shown. Three AE appraisal subjects served as pre-board exams for BS ABE students.

Data Collection Tools

In this study, the proponents used WEKA classification software to create a prediction model for the study. According to Hall et al. (2009), WEKA is an open-source software package offering data preparation, clustering, machine learning algorithm implementation, and visualization features for data analysts to develop machine learning models using various classifiers. The proponents find the best classifier in terms of accuracy.

Research Procedures

The proponents gathered all the necessary documents from the database of Cavite State University. These documents contained two databases: the raw grades of students who took the BSABE exam and the PRC list of exam takers of CvSU. The proponents cleaned the data by removing all insufficient grades from the enrollee. The researchers also implemented data sorting to collect all the numerical grades for the first-time taker based on the PRC list. The proponents, however, entered grades into Microsoft Excel without sensitive data such as the name of the student and their number. It was done to keep the information confidential and not to share it with anyone. Officials from the university handled all concerns related to student grades and admission standards. They also served as quality control in order to ensure that the assessment of student academic performance was objective. The dataset created was sent to a co-author for evaluation, testing, and verification. The proponents converted datasets from comma separated values or CSV to attribute-relation files ARFF using WEKA software. Once the data was converted, the information was entered into WEKA. WEKA then ranked subjects based on their chances of passing board licensing examinations.

Ethical Considerations

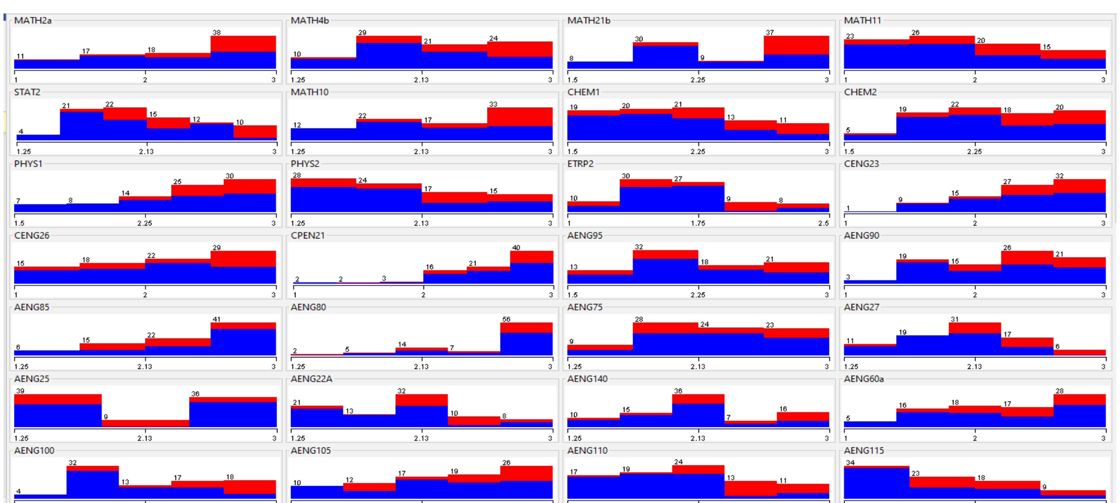
Ethics are of the utmost importance in any research project that includes human or animal subjects. For this research, the authors employed a quantitative, quasi-experimental design in order to explore any correlations between high subject scores and passing the licensure examination.

As for ethical considerations, the authors did not include an in-depth section detailing their protocols in this study based on what information was made available to them. However, they obtained all required documents from Cavite State University's administration, such as two databases of student scores who took BSABE exams and the PRC list of exam takers at Cavite State University. The proponents cleaned this data by eliminating insufficient grades from enrollees before entering it into Microsoft Excel without including sensitive student details like names or numbers in order to maintain the privacy of data and not reveal this sensitive data to anyone outside. University officials handled any concerns pertaining to student grades or admission standards and served as quality control to ensure an unbiased assessment of academic performance by students. Furthermore, the authors excluded graduates who had not yet taken their licensure examination or those making another attempt at it.

Results and Discussion

In this study, the proponents set the ARFF format in WEKA software. Once the dataset was converted, the proponents explored the graphical representation of the file in the preprocessing section. Then, they simulated all of the available classifiers using the classify menu and find the best classifier in terms of accuracy in 5-fold cross-validation. In the case of the same highest accuracy of the classifier, the proponents determined the fastest simulation time of the classifier. Then the best classifier was chosen.

Figure 2 visually displays the subjects and their classes. The positive class was represented by the color red, while the negative class was represented by the color blue. The assignment of colors to categorical values was done automatically. These visualizations help readers understand the distribution and classes of the subjects at a glance, without requiring in-depth knowledge of the technical terms used.



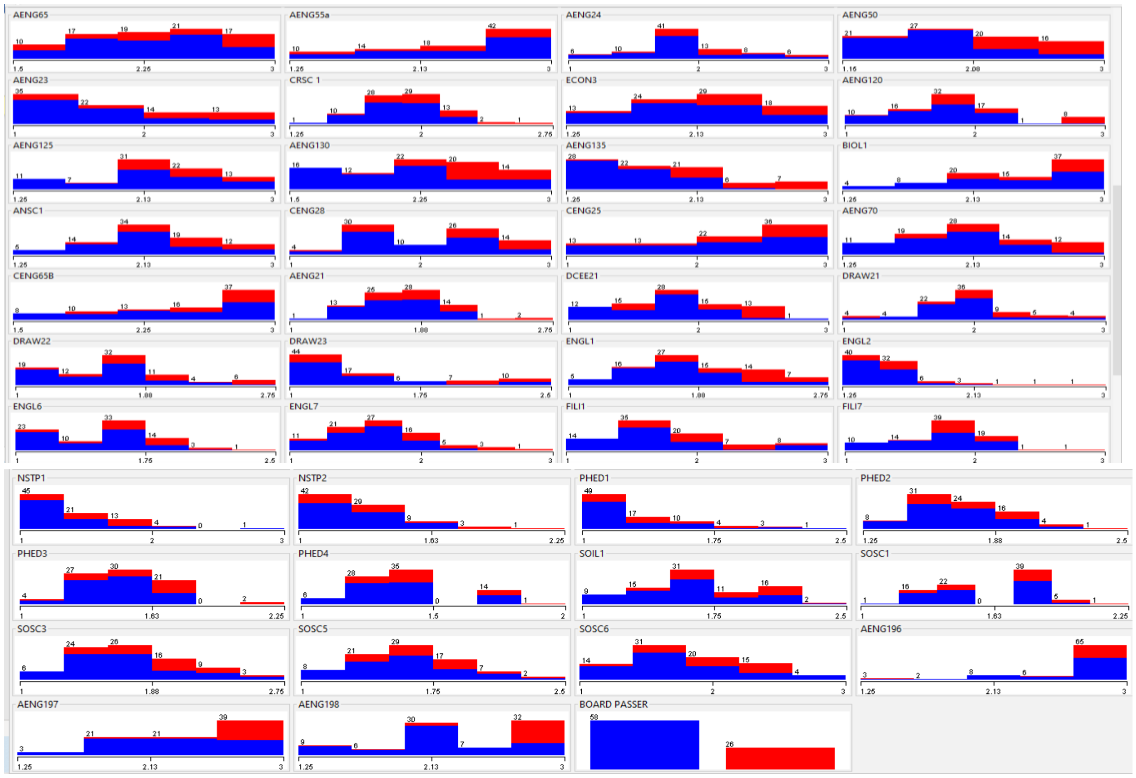


Figure 2. *Visual Representations of BS ABE Undergraduate Grades*

Weka Classifier Simulation Result

Table 1 displays the simulation results of five-fold cross-validation using various classifiers in WEKA. The highest accuracy rate of 92.8571% was achieved by the IBK Classifier, which had an operating time of 0.0 second. The Random Forest Classifier also performed well, with an accuracy rate of 91.667% and an operating time of 0.01 second. These results indicate that both classifiers are effective in accurately classifying the subjects, with the IBK Classifier showing superior accuracy. The table provides valuable information about the performance and efficiency of each classifier.

Table 1. The Total Accuracy and Estimated Time of the Used Classifier in WEKA

Categorized Classifier	Sub -Classifier	Accuracy	Training time
Bayers Classifier	BayersNet	86.9048%	0 second
	NaiveBayers	82.1429%	0 second
	NaiveBayersMultinomial	72.619%	0.01 seconds
	NaiveBayersMultinomialText	69.0476%	0 second
	NaiveBayersMultinomialUpdatable	72.619%	0.01 seconds
	NaiveBayersUpdatable	82.1429%	0 second
Functions Classifier	Logistics	77.381%	0.04 seconds
	MultilayerPerception	90.4762%	5.63 seconds
	SGD	86.9048%	0.11seconds
	SGDText	69.0476%	0.07 seconds
	SimpleLogistics	84.5238%	0.39 seconds
	SMO	86.9048%	0.07 seconds
	VotedPerception	69.0476%	0.02 seconds

Lazy Classifier	IBK	92.8571%	0 second
	KStar	88.0952%	0.12 seconds
	LWL	70.2381%	0 seconds
Meta Classifier	AdaboostM1	88.0952%	0.03 seconds
	AttributeSelectedClassifier	82.1429%	0.13 seconds
	Bagging	82.1429%	0.06 seconds
	ClassificationViaRegression	77.381%	0.26 seconds
	CVParameterSelection	69.0476%	0.02 seconds
	FilteredClassifier	80.9524%	0.02 seconds
	IlerativeClassifierOptimizer	88.0952%	0.54 seconds
	LogitBoost	90.4762%	0.2 seconds
	MultiClassClassifier	77.381%	0.05 seconds
	MultiClassClassifierUpdatable	86.9048%	0.03 seconds
	MultiScheme	69.0476%	0 second
	RandomComittee	85.7143%	0.02 seconds
	RandomizableFilteredClassifier	84.5238%	0.02 seconds
	RandomSubSpace	80.9524%	0.06 seconds
	Stacking	69.0476%	0 second
	Vote	69.0476%	0 second
	WeightInstancesHandlerWrapper	69.0476%	0 second
Misc Classifier	InputMappedClassifier	69.0476%	0 second
Rules Classifier	DecisionTable	79.7619%	0.14 seconds
	JRip	71.4286%	0.07 seconds
	OneR	71.4286%	0 second
	PART	82.1429%	0.03 seconds
	ZeroR	69.0476%	0 second
Trees Classifier	DecisionStump	69.0476%	0 second
	HoeffdingTree	82.1429%	0.31 seconds
	J48	80.9524%	0.02 seconds
	LMT	84.5238%	0.4 seconds
	RandomForest	91.6667%	0.1 seconds
	RandomTree	77.381%	0 second
	REPtree	76.8571%	0.1 seconds

IBK Classifier Simulation Result

Figure 4 shows the actual simulation result of the five-fold cross-validation of the IBK classifier. It shows the actual results of the stratified cross, detailed accuracy of class, and confusion matrix of the IBK classifier. The stratified cross-validation results demonstrate how well the IBK classifier performed in classifying the instances. It presents the accuracy rates and other performance metrics that indicate the effectiveness of the classifier in correctly predicting the class labels.

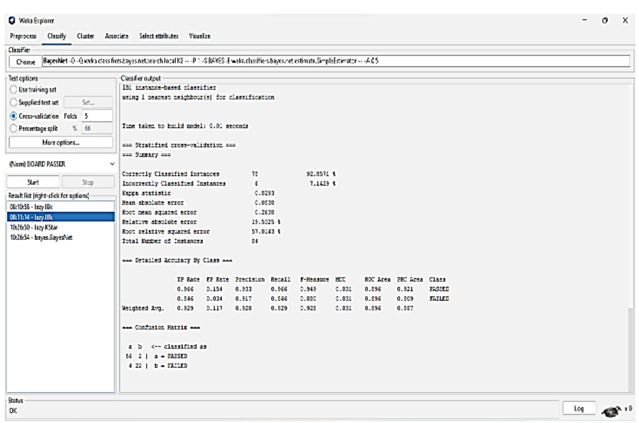


Figure 4. *Validation of IBK Classifier Using 5 -Fold Cross Validation*

The detailed accuracy by class provides a more specific breakdown of the classifier's performance for each individual class. It shows metrics such as precision, recall, F-measure, MCC, ROC area, and PRC area. These metrics help assess the classifier's ability to correctly classify instances belonging to different classes. The confusion matrix in Figure 4 offers a visual summary of the classification results. It provides a matrix that shows the number of instances classified correctly and incorrectly for each class. This matrix helps evaluate the performance of the classifier by identifying any patterns of misclassification or confusion between classes.

In summary, Figure 4 presents a comprehensive overview of the simulation results obtained from the five-fold cross-validation using the IBK classifier. It includes information about the accuracy, detailed accuracy by class, and the confusion matrix. These visualizations provide insights into the performance and effectiveness of the IBK classifier in accurately classifying the instances. Regarding the IBK classifier, it stands for Instance-Based Learning with k-nearest Neighbors. It is a type of supervised machine learning algorithm that classifies instances based on their similarity to other instances in the training data. The IBK classifier determines the class label of a new instance by comparing it to its k nearest neighbors in the feature space. The class label of the majority of the nearest neighbors is assigned to the new instance. In short, the IBK classifier utilizes the concept of nearest neighbors to make predictions and is commonly used for classification tasks in various domains.

Table 2 presents the results of the stratified cross-validation performed on the dataset using the IBK classifier in WEKA. The table provides important metrics and statistics to evaluate the performance of the classifier. The "Correctly Classified Instances" column indicates that 78 instances were correctly classified, accounting for an accuracy rate of 92.8571%. This means that the classifier accurately predicted the class labels for the majority of the instances in the dataset.

On the other hand, the "Incorrectly Classified Instances" column shows that six (6) instances were classified incorrectly, resulting in an error rate of 7.1429%. These instances were misclassified by the classifier. The "Kappa statistic" is a measure of agreement between the predicted and actual classifications. In this case, the Kappa statistic is calculated to be 0.8293, indicating a substantial level of agreement beyond what can be attributed to chance. The "Mean absolute error" and "Root mean squared error" provide measures of the prediction errors made by the classifier. The values reported in the table are 0.0838 and 0.2638, respectively. These values represent the average absolute and squared differences between the predicted and actual class labels. Lower values indicate better accuracy and precision of the classifier. The "Relative absolute error" and "Root relative squared error" provide measures of the prediction errors relative to the magnitude of the actual values. These values are reported as 19.5025% and 57.0143%, respectively. These measures can help assess the relative accuracy and precision of the classifier's predictions. Lastly, the table shows that the dataset used for cross-validation contains a total of 84 instances.

Table 3. Detailed Accuracy by Class Using WEKA

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.966	0.154	0.933	0.966	0.949	0.831	0.896	0.921	PASSED
	0.848	0.034	0.917	0.845	0.880	0.831	0.896	0.809	FAILED
Weighted Avg.	0.929	0.117	0.928	0.929	0.928	0.831	0.896	0.887	

Table 3 shows the detailed accuracy by class using WEKA with IBK classifier. As seen in the table below, the weighted average is measured at 0.929 in the rate of true positives, 0.928 in precision, and a low 0.117 in the rate of false positives, indicating that the dataset is accurate. It is also shown on Table 4 the weighted average of recall, F-measure, MCC, ROC area, and PRC area of the dataset being used.

Table 4: The Confusion Matrix of BOARD PASSERS Using IBK Classifier

Total	Passed	Failed
58	56	2
26	4	22

Table 4 shows the confusion matrix of class board passers using the IBK classifier. The "passed" variable shows that 56 instances were passed and 2 instances were predicted to fail in the licensure examination, and the failed variable shows that 22 instances failed and 4 failed results were predicted to pass the licensure examination. The results of the IBK classifier in the data produced a 92.8571% accuracy rate.

Table 5. Top 20 Subjects of BSABE Undergraduate Grade that Affect the Board Licensure Performance

Rank	Subject Codes	Subject Description	Subject category	Average Merit	Average Rank
1	AENG110	Soil And Water Conservation Engineering	Major Subjects	0.112 +- 0.03	5 +- 3.29
2	MATH21B	Analytic Geometry	Engineering Mathematics and science subjects	0.114 +- 0.019	5.2 +- 3.19
3	ENGL1	Study And Thinking Skills In English	General Subjects	0.107 +- 0.032	5.4 +- 3.61
4	AENG198	Ae Competency Appraisal 3	Engineering Appraisal	0.111 +- 0.03	6.6 +- 7.31
5	IENG90	Introduction To Operations Research	Major Subjects	0.095 +- 0.027	7 +- 5.62
6	AENG70	Agricultural Power And Energy Resource	Major Subjects	0.089 +- 0.023	7.6 +- 4.45
7	MATH10	Differential Calculus	Engineering Mathematics and science subjects	0.076 +- 0.024	10.2 +- 7.44
8	AENG50	Principles Of Fishery Science	Major Subjects	0.08 +- 0.026	10.4 +- 5.61
9	ETRP2	Agricultural Entrepreneurship And Management	Major Subjects	0.078 +- 0.011	10.6 +- 3.83
10	AENG135	Agricultural Waste Management	Major Subjects	0.077 +- 0.029	11.6 +- 5.46
11	DCEE21	Basic Electricity And Electronics	Major Subjects	0.077 +- 0.036	12.2 +- 8.75
12	AENG100	Soil And Water Conservation Engineering	Major Subjects	0.066 +- 0.028	14.6 +- 7.61
13	AENG130	Aquaculture Engineering	Major Subjects	0.065 +- 0.022	14.6 +- 6.74
14	AENG25	Materials Of Engineering	Major Subjects	0.052 +- 0.021	18.8 +- 9.15
15	DRAW23	Computer-Aided Drafting And Design (CADD)	Major Subjects	0.054 +- 0.042	18.8 +-12.69
16	AENG22A	Computer Application In Engineering 1	Major Subjects	0.04 +- 0.023	22.6 +- 9.71
17	CHEM1	General Chemistry 1	Engineering Mathematics and science subjects	0.041 +- 0.021	22.8 +-10.15
18	CENG26	Engineering Management	Major Subjects	0.038 +- 0.024	23.2 +- 9.2
19	AENG140	Agricultural Eng'g Law And Professional Ethics	Major Subjects	0.037 +- 0.033	24.4 +-14.55
20	STAT2	Probability And Statistics	Engineering Mathematics and science subjects	0.039 +- 0.026	24.8 +-11.03

Table 5 shows the prediction model of the BSABE subjects using the Select Attributes menu of the WEKA Explorer. It verifies the rank of the subjects using five-fold cross-validation. As seen on the table, it presents the top 20 subjects that affect the performance in passing the Licensure Examination of BSABE.

Based on these results, Soil and Water Conservation Engineering or AENG110 has the highest rank among all subjects of the BS ABE program with 0.112 +- 0.03 average merit 5 +- 3.29 average rank. It was followed by Analytic Geometry or MATH 21A which has an average merit of 0.114 +- 0.019 and an average rank of 5.2 +- 3.19, and Study and Thinking Skills in English or ENGL1 which has an average merit of 0.114 +- 0.019 and average rank of 5.2 +- 3.19.

In this table, most of the top subjects of the BS ABE Program belong to the major subjects indicating the critical point subjects in passing the board licensure examination. This includes agricultural mechanization and equipment allied subjects, soil and water resources development and conservation allied subjects, and rural electrification agricultural processing and agricultural structures allied subjects. This category must be known by the BSABE students to prepare for the Licensure Examination.

In addition, the engineering mathematics and science subjects were also among the priority subjects to pass the engineering board licensure examination. Subjects including MATH21B, MATH10, CHEM1, and STAT2 were among the top 20 subjects initialized by WEKA. These subjects serve as the fundamentals in applying lessons in major subjects including crops science subjects, electrification allied subjects, and research-related subjects. Furthermore, the AE Competency Appraisal is among the most important subjects in which ABE students which serves as the pre-board exam preparation. The AE Competency Appraisal 3 was included in the table which has an average merit of 0.111 +- 0.03.

Lastly, the general academic subjects particularly English subjects were also included in the predicted grades to pass the board licensure examination. The findings support the claim of Zhang et al., (2013) and Simbulas et al., (2015) which revealed a clear significant link between linguistic competencies, such as verbal reasoning and problem-solving abilities (Batuctoc, n.d.; Simbulas et al., 2015). Moreover, because English is the medium of instruction, students who were more fluent in the language performed better in writing, speaking, grasping, and understanding the instructions and lessons provided to them in professional courses (Oducado et al., 2020). ENGL1 serves as the foundation for improving the English reading comprehension and critical thinking of the students of the BSABE program. As a result, the proponents proved that takers with high scores in Engineering, Mathematics, and Science subjects, general appraisal, major subjects, and English general subjects had a high chance of passing the licensure examination.

Conclusion and Future Works

This study aimed to determine whether the grades in more difficult subjects correlate with the passing rate on the licensure examination. It also determined the most accurate and influential classifier by using WEKA Classification Software. According to these findings, IBK was the most accurate classifier, with an accuracy of 92.8571%. The training time for the IBK Classifier was 0 second. The three most influential subjects were AENG110 (MATH21B), ENGL1, and MATH21B. These subjects had a significant impact on the subject's ability to pass the board exam. The proponents concluded that there was a correlation between a higher grade in subjects in the BSABE Program and passing the licensure examination.

Based on the findings, if the first-time examinee gets a high grade in Engineering math and sciences, competency assessment subjects, majors, and English subjects during their undergraduate year, there is a greater chance of passing the licensure examination. The top engineering subjects for passing the licensure examination were Calculus, Physics, and electrical allied fields. Due to external factors, the researchers were only able to use 84 examples in this study. The results could be different if a larger dataset would be used. This study may be continued by future researchers who would use a larger dataset. This would result in a more accurate and relevant engineering solution using data mining. This study would help to understand the trends of licensure performance for the BS Agricultural and Biosystems Engineering Students to ensure they pass their respective board exams. It also helps the faculty and administrators of the college to better guide their students to achieve higher scores in the board examination.

Future work should include exploring other data mining techniques that might produce superior results such as decision trees or neural networks. Such advanced approaches include sophisticated analyses, while simultaneously helping uncover intricate relationships among variables. Other future considerations may also include conducting a follow-up study that evaluates the effectiveness of strategies designed to improve licensure exam performance, such as review classes or mentoring programs, formal review sessions, and the scores of their pre-board exams (if any) in different topics covered. The development of a prediction model for undergraduate grade and licensure examination performance in the Bachelor of Science in Agriculture and Biosystems Engineering had significant implications for students, educational institutions, and the field as a whole. This study's findings can convey helpful insights that can enhance academic results and guide decision-making.

In addition, there are some implications and suggestions derived from the study. First, the prediction model can assist in identifying the subjects or areas of study that have a strong correlation with success on the licensure exam. By analyzing the results of the model, students can prioritize their exam preparation and devote more time and effort to the most important subjects. In addition, educational institutions can use this data to improve their curricula, ensuring that these essential subjects receive the necessary attention and resources.

Second, using the prediction model, students can customize their exam preparation strategies in accordance with their individual strengths and deficiencies. By understanding which subjects have a greater impact on their performance, students can concentrate in enhancing their knowledge and skills in those areas. They can utilize additional study materials, tutoring, and specialized exam preparation courses to improve their knowledge and abilities.

Third, the study's findings can assist educational institutions in evaluating and enhancing their curriculum. If certain subjects consistently demonstrate a strong correlation with licensure exam success, institutions may consider enhancing the subject's content, teaching methods, or resources. This can ensure that students are well-prepared and have a greater chance of performing well in the licensure exam.

Fourth, institutions can create support programs to aid students in areas designated by the prediction model as crucial. These programs may consist of seminars, study groups, or mentoring initiatives that provide students with additional guidance and resources to excel in the identified subjects. By providing targeted assistance, institutions can contribute to improved student performance and exam pass rates.

Lastly, the prediction model can be used as a tool for monitoring and evaluating educational programs on an ongoing basis. By routinely analyzing data and revising the model, institutions can evaluate the efficacy of any modifications to the curriculum or support programs. This iterative process can help identify areas for development and ensure that educational objectives and intended outcomes remain aligned.

References

[1] Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics, 37*, 13–49. https://doi.org/10.1016/j.tele.2019.01.007

[2] Alkharusi, H., & Al-Hinai, A. (2018). Predicting students' academic performance using their demographic and academic data: A case study. *International Journal of Information and Education Technology, 8*(4), 268-273.

[3] Antonio, J., Malvar, R., Ferrer, M., & Pambuena, E. (2016). Licensure examination for teachers results from 2010 to 2013 of PUP San Pedro's Bachelor in Secondary Education major in Mathematics and English graduates and its relationship on their academic performance. *Asia Pacific Journal of Multidisciplinary Research, 4*(44), 17-22. http://www.apjmr.com/wp-content/uploads/2016/09/AP JMR-2016.4.4.03.pdf

[4] Bachhal, P., Ahuja, S., & Gargrish, S. (2021). Educational Data Mining: A review. *Journal of Physics Conference Series, 1950*(1). https://doi.org/10.1088/1742-6596/1950/1/012022

[5] Bae, R.E, Arboleda, E., Andilab, A., Dellosa, R. M., (2019). Implementation of template matching, fuzzy logic and k nearest neighbor classifier on Philippine banknote recognition system. *International Journal of Scientific and Technological Research, 8* (8), 1451-1453.

[6] Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning, 4*(2), Article 17. https://doi.org/10.20429/ijsotl.2010.040217

[7] Banluta, J. (2013). *Relationship of the academic rating and board examination performance of the electronics engineering graduates*. https://www.researchgate.net/publication/324416083_Relationship_of_the_Academic_Rating_and_Board_Examination_Performance_of_the_Electronics_Engineering_Graduates

[8] Batuctoc, L.V. (n.d.). Effectiveness of metacognition-based reading enrichment program to students' reading comprehension. *International Journal of Advanced Research, 9*(2), 111-135. https://doi.org/10.21474/IJAR01/12425

[9] Bruce, S., Crawford, E., Wilkerson, G., Dale, R. B., Harris, M., & Rausch, D. (2019). Prediction modeling for board of certification exam success for a professional master's athletic training program. *Journal of Sports Medicine and Allied Health Sciences: Official Journal of the Ohio Athletic Trainers' Association, 5*(2). https://doi.org/10.25035/jsmahs.05.02.07

[10] Chan-Rabanal, G. (2016). Academic achievement and LET performance of the Bachelor of Elementary Education graduates, University of Northern Philippines. *International Journal of Scientific and Research Publications, 6*(6), 455–461.http://www.ijsrp.org/research-paper-0616.php?rp=P545511

[11] CEIT News and Updates – Cavite State University. (2022).Cvsu.edu.ph.

[12] Dagdag, J. (2018). Predictors of performance in the Licensure Examination for Agriculturists: Bases for a proposed plan of action. *Asia Pacific Journal of Multidisciplinary Research,6*(2), 113–120. http://www.apjmr.com/wp-content/uploads/2018/02/APJMR2017.6.2.14.pdf

[13] Ferrer, F. (2016). Performance in the Engineer Licensure Examinations: Philippines, 2011-2015. *International Journal of Advance Research in Science and Engineering, 5*(1), 5–58. http://www.ijarse.com/images/fullpdf/1454832385_238 S.pdf

[14] Garcia, J.A., Arboleda,E., & Galas, E., (2020). Identification of visually similar vegetable seeds using image processing and fuzzy logic. *International Journal of Scientific and Technology Research, 9*(2),4925-28.

[15] Geollegue, K. W., Arboleda, E., Dizon, A. A. (2022). Seed of rice plant classification using coarse tree classifier. *International Journal of Artificial Intelligence, 11*(2), 727~735. https://doi.org/10.11591/ijai.v11.i2.pp727-735

[16] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

[17] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter, 11*(1), 10-18. https://dl.acm.org/doi/10.1145/1656274.1656278

[18] Ladia, M., & Nool, N. (2020). Performance in the Licensure Examination for Teachers among the graduates of Isabela State University, Echague, Isabela, Philippines. *Journal of Critical Reviews, 7*(11). https://www.academia.edu/7963956/Performance_in_the_Licensure_Examination_for_Teachers_of_State_Uni versities_and_Colleges_in_Region_III

[19] Landrum, R. E., & Harrold, R. L. (2015). Predicting performance on the counselor preparation comprehensive examination from undergraduate academic performance. *Counselor Education and Supervision, 54*(3), 154-166.

[20] Laguador, J., & Dizon, N. (2013). Academic achievement in the learning domains and performance in Licensure Examination for Engineers among LPU's mechanical and electronics engineering graduate. *International Journal of Management, IT and Engineering, 3*(8), 347-378. https://research.lpubatangas.edu.ph/wp-content/uploads/2014/04/IJMIE-Academic-Achievement-in-the-Learning.pdf

[21] Maaliw, R. R. (2021). Early prediction of Electronics Engineering Licensure Examination Performance using random forest. *In 2021 IEEE World AI IoT Congress, AIIoT 2021* (pp. 41–47). https://doi.org/10.1109/AIIoT52608.2021.9454213

[22] Manalo, V. M., Arboleda, E. R., Dioses, J.L., & Dellosa, R.M., (2019). Differentiation among lettuce (L. Sativa) seed varieties grown in gourmet farms, Silang Cavite, Philippines using image processing with fuzzy logic and KNN as classifiers. *International Journal of Scientific and Technology Research, 8*(10), 318-321 https://scholar.google.com/scholar?oi=bibs&cluster=11977495805413999811&btnI=1&hl=en.

[23] Nicolas, H. J., De Guzman, J., Tejada, R. C., & Capalad, R. (2020). Student determinants in the Licensure Examination for Agriculturists of a state college in the Philippines. *International Journal of Educational Sciences, 28*(1-3). https://doi.org/10.31901/24566322.2020/28.1-3.1117

[24] Oducado, R. M. F., Sotelo, M. G., Ramirez, L. M. M., Habaña, M. P., & Belo-Delariarte, R. G. (2020). English language proficiency and its relationship with academic performance and the Nurse Licensure Examination. *Nurse Media Journal of Nursing, 10*(1), 46-56. https://doi.org/10.14710/nmjn.v10i1.28564

[25] Pau, A., Jeevaratnam, K., Chen, Y. S., & Fallis, D. (2017). The impact of study habits on academic performance: A study of students in a pharmacy program. *Currents in Pharmacy Teaching and Learning, 9*(2), 199-206.

[26] Rabe K.E., Arboleda, E., Andilab, A., & Dellosa, R. (2019). Fuzzy logic based vehicular congestion estimation monitoring system using image processing and KNN classifier. *International Journal of Innovative Technology and Exploring Engineering, 8*(12), 2278-3075.

[27] Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353-387.

[28] Rienecker, L., & Stray Jørgensen, P. (2019). *Studying at university: A guide for first-year students (2nd Edition).* Gyldendal Akademisk.

[29] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. *In Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318-362). MIT Press.

[30] Rustia, R., Cruz, M. M., Burac, M. A., & Palaoag, T. (2018). Predicting student's board examination performance using classification algorithms. *In ACM International Conference Proceeding Series* (pp. 233–237). https://doi.org/10.1145/3185089.3185101

[31] Sánchez-Martín, J., Suárez-Cantero, C., & Delgado-Rodríguez, M. (2019). Influence of academic performance and academic self-efficacy on the preparation for university entrance exams: A study among Spanish students. *Higher Education Research & Development, 38*(6), 1195-1209.

[32] Sverdlik, A., & Hall, N. C. (2016). Motivation and self-regulated learning: A multilevel analysis. *Learning and Individual Differences, 45*, 161-168.

[33] Simbulas, L., Regidor, B., & Catulpos, R. (2015). Reading comprehension and mathematical problem solving skills of University of the Immaculate Conception freshmen students. *UIC Research Journal, 21*(2), 67-70. https://ejournals.ph/article.php?id=12835

[34] Swaen, B. (2019, May 4). *Developing a conceptual framework for research example*. Scribbr. https://www.scribbr.com

[35] Su, Y.-S., & Lai, C.-F. (2021). Applying educational data mining to explore viewing behaviors and performance with flipped classrooms on the social media platform Facebook. *Frontiers in Psychology, 12*, 1-8. https://doi.org/10.3389/fpsyg.2021.653018

[36] Tamayo, A., Bernardo, G., & Eguia, R. (2014). Readiness for the licensure exam of the engineering students. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.2395037

[37] Tarun, I. (2017). Prediction models for licensure examination performance using data mining classifiers for online test and decision support system. *Asia Pacific Journal of Multidisciplinary Research, 5*(3), 10–21. http://www.apjmr.com/wp-content/uploads/2017/06/APJMR-2017.5.3.02.pdf

[38] Zeidner, M., & Klug, H. J. (2013). The role of study time and study strategy in predicting academic achievement outcomes: A review. *Educational Psychology Review, 25*(3), 375-405.

[39] Zhang, X., Koponen, T., Räsänen, P., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2013). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Development, 85*(3), 1091–1107. https://doi.org/10.1111/cdev.12173

Acknowledgment

The researchers would like to express their sincerest thanks to God for blessing them with strength and resilience to meet the many obstacles of daily academic work life. Furthermore, they appreciate and recognize Dr. Edwin R. Arboleda as co-author and professor of Methods of Research, for his invaluable assistance and mentorship which led to the creation of high-quality research studie using reliable programs for data analysis, and for encouraging writing research within their class. They give thanks to the Cavite State University Registrar's Office for giving them access to its reliable resources, such as its student database and PRC verification board exam database, both of which played an invaluable role in their research endeavor.

The researchers also recognize their classmates for providing reliable literature and helping them to hone their grammar skills in research. Their contributions had made this research study possible and should not go unappreciated.