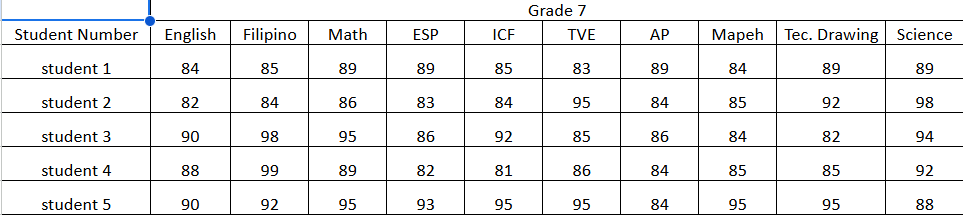
**CHAPTER IV**

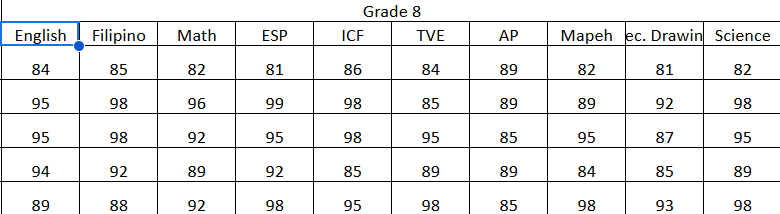
**RESULTS AND DISCUSSIONS**

## **4.1 Results**

This chapter presents the results obtained from the collection, preprocessing, and analysis of data from Butuan City School of Arts and Trades (BCSAT) junior high school students, as well as the performance of the multinomial logistic regression model in predicting senior high school strand preferences.

## **4.1.1 Data Collection**





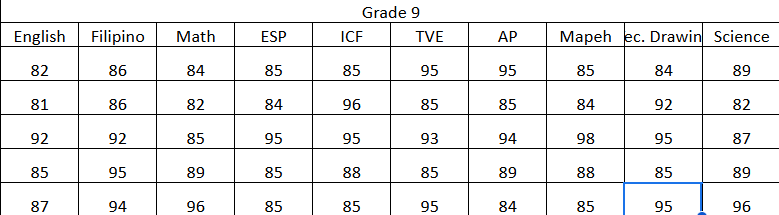
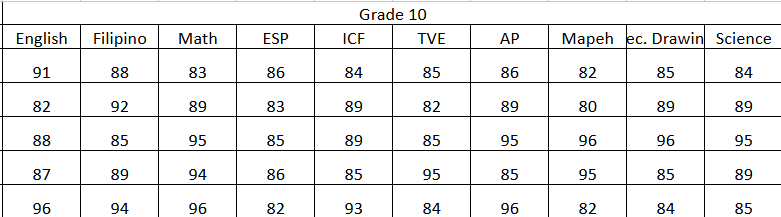
  


Figure 5: Sample of Collected Student Data

The data collection process involved gathering academic records of junior high school students from BCSAT. The researchers obtained permission from the school principal to access the 120 student grade records From there Grade 7 to Grade 10 for their General average per subject from the school year of 2024 -2025,out of 7 sections in Grade 10 only 4 sections that we gather the data while ensuring student privacy by excluding their names. The collected data included grades in various subjects such as Mathematics, Science, English, Filipino, TVEL, ICF,Mapeh, ESP, and Araling Panlipunan,Technical Drawing .

### **4.1.2 Pre-Processing**

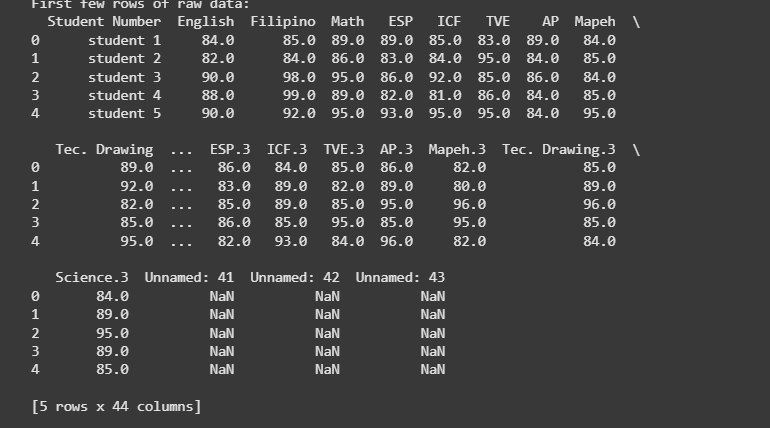
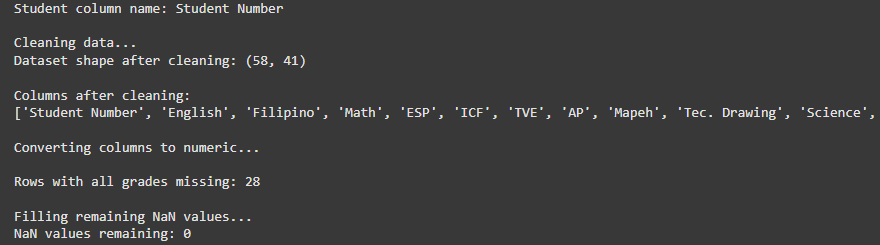


Figure 6. Loading dataset

The Student Grades dataset contains academic performance data across multiple subjects. It includes attributes such as Student Number, scores in core subjects like English, Math, Science, Filipino and specialized subjects such as Technical Drawing, ICF, MAPEH, AP, ESP (Edukasyon sa Pagpapakatao) and TVEL. The dataset also contains some columns with missing or unnamed values. Using mean imputation helps to fill in these missing values by replacing them with the average score from Grade 7 to Grade 9, preserving the dataset's integrity and ensuring no loss of valuable student performance insights.

**4.1.3 Data Cleaning**

 Figure 7. Data Cleaning Results

The data cleaning process involves removing unnecessary columns (e.g., unnamed columns), converting grade columns to numeric format, and identifying rows with missing values. Rows where all grade columns are missing are removed to maintain data integrity. This step ensures the dataset is clean, consistent, and ready for accurate model training.

After cleaning, the dataset contains 120 rows and 12 rows. All remaining missing values have been successfully filled, ensuring no empty values in the dataset.

**4.1.4 Feature Selection**



Figure 8. Features Selection

This figure shows the feature selection process for the student grades dataset. The features (**X**) include various subjects like **English, Math, Science**, and others. The target variable (**y**) **preferred strand** represents the academic outcome.

**4.1.5 Feature Engineering**

| **STRANDS** | **SUBJECTS/FEATURES** |
| --- | --- |
| STEM | ['Math', 'Science'] |
| ABM | ['Math', 'English'] |
| HUMMS | ['English', 'Filipino', 'AP'] |
| GAS | ['English', 'Math', 'Science', 'Filipino', 'AP'] |
| TVEL | ['TVE', 'ICF', 'Tec. Drawing'] |

Table 5: Strand Features

Table no 5. Shows the defined subject group for each strand. These subjects are the features that are used in training the Multinomial Logistic Model. For example, if a student’s average grade in Math and Science is greater than other subjects the model will analyze these features and predict or suggest the ‘STEM’ strand for the student. In order for the model to predict, it will find the highest average (ex. Math grade, Science grade).

**4.1.5 Train and Test Split**

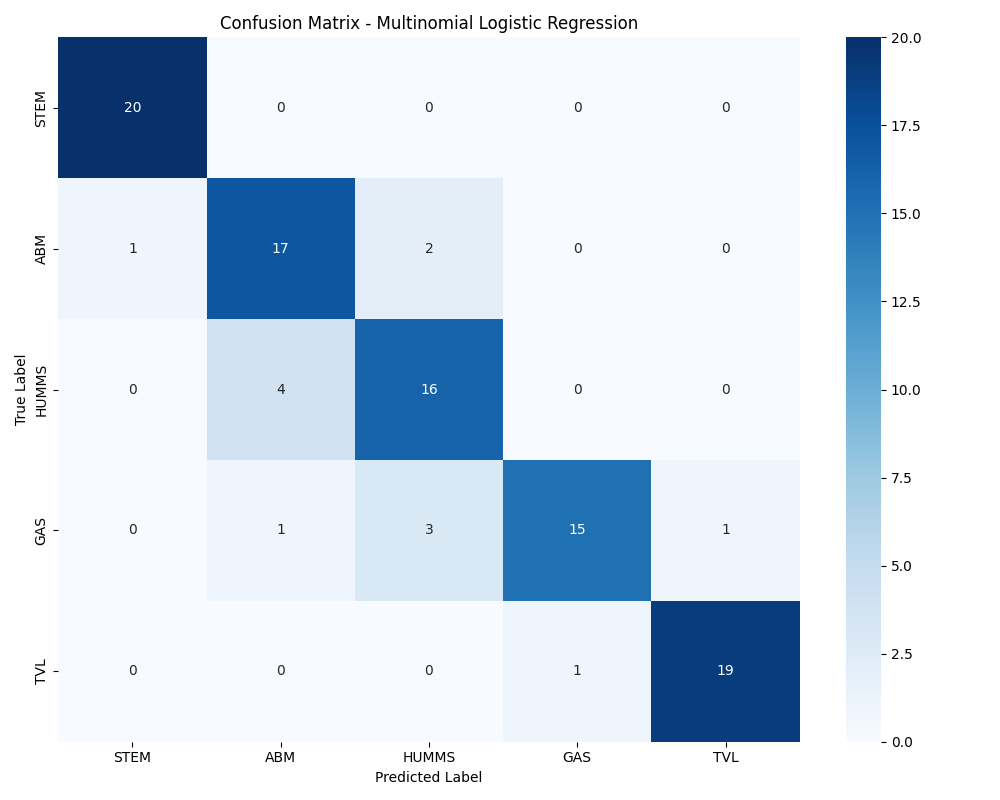
| **Data Train Size** | **Data Test Size** |
| --- | --- |
| **70%** | **30%** |

**Figure 9. Data Split**

The dataset is now ready for training and testing, the researcher set the train-test size for the purpose of model evaluation in the later process. Train-test split process divides the dataset into training and testing subsets. Here, 70% of the data is used for training the model, while 30% is reserved for evaluating its performance on unseen data.

**4.1.6 Model Training Result**

**4.1.6.1 Multinomial Logistic Regression Confusion Matrix**

****

**Figure 10**. Multinomial Logistic Regression Confusion Matrix

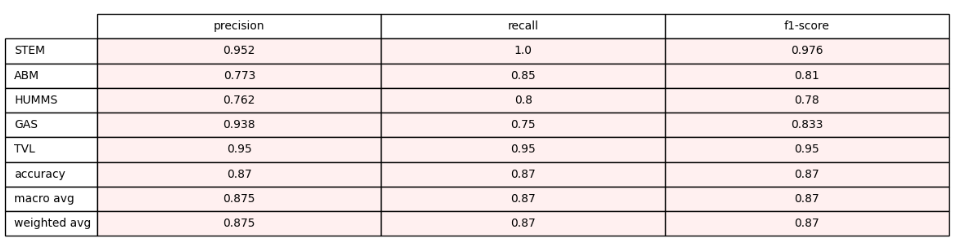
The figure above shows the confusion **matrix** for a **multinomial logistic regression** model. It shows how well the model predicts categories (labels) by comparing the **true labels** (what the actual categories are) with the **predicted labels** (what the model guessed). The categories here are STEM, ABM, HUMMS, GAS, and TVL likely representing different academic tracks or strands.

* **Rows** (True Labels): These are the actual categories (e.g., STEM, ABM, etc.).
* **Columns** (Predicted Labels): These are the categories the model predicted.
* **Numbers in the cells**: They show how many times the model predicted a certain category for a true category.
* **Diagonal (top-left to bottom-right)**: Correct predictions (e.g., true STEM predicted as STEM).
* **Off-diagonal**: Incorrect predictions (e.g., true STEM predicted as ABM).

**Analysis:**

* **STEM**: 20 were correctly predicted as STEM (good!). None were misclassified.
* **ABM**: 17 were correctly predicted as ABM, but 1 was predicted as STEM, and 2 as HUMMS.
* **HUMMS**: 16 were correctly predicted as HUMMS, but 4 were predicted as ABM.
* **GAS**: 15 were correctly predicted as GAS, but 1 was predicted as ABM, 3 as HUMMS, and 1 as TVL.
* **TVL**: 19 were correctly predicted as TVL, but 1 was predicted as GAS.

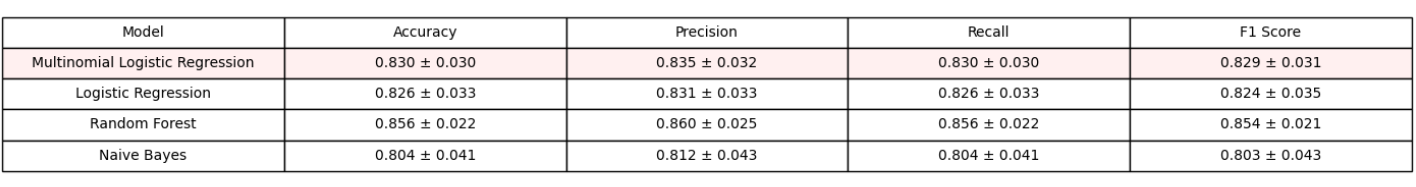
**4.1.6.2 Multinomial Logistic Regression Classification Report**



**Figure 11.** Multinomial Classification Report

The figure above shows the classification report for a multinomial logistic regression model. It shows how well it predicts five categories (STEM, ABM, HUMMS, GAS, TVL). It uses three metrics: **precision** (how many predicted items were correct), **recall** (how many actual items were correctly predicted), and **f1-score** (a balance of precision and recall). STEM and TVL performed best with high scores (e.g., STEM: 0.952 precision, 1.0 recall), while HUMMS had the lowest (0.762 precision, 0.8 recall). The overall **accuracy** is 0.87 (87% correct predictions). The **macro avg** (average across categories) and **weighted avg** (adjusted for category size) both show 0.875 for precision, 0.87 for recall, and 0.87 for f1-score, meaning the model is fairly balanced but struggles slightly with HUMMS and ABM.

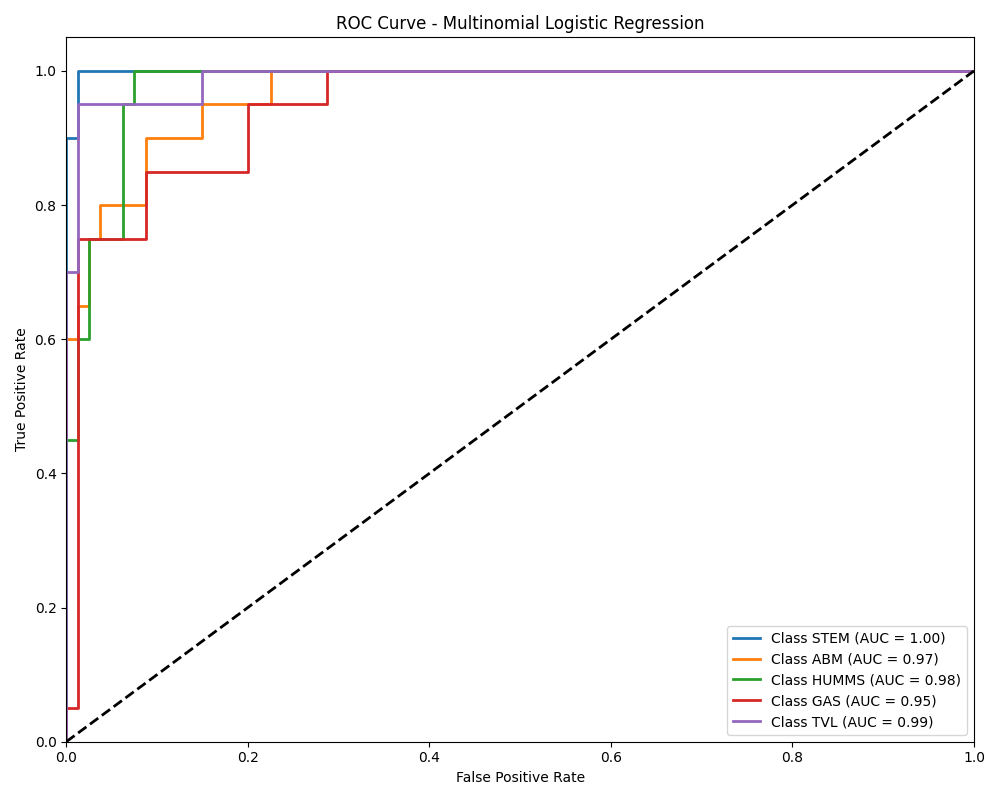
**4.1.6.3 Multinomial Cross-Validation Report**

****

**Figure 12.** Multinomial Cross-Validation Report

This figure above shows the cross-validation results for four models; Multinomial Logistic Regression, Logistic Regression, Random Forest, and Naive Bayes, measuring their performance in accuracy, precision, recall, and F1-score, with standard deviations to show consistency. Random Forest performs best overall, with the highest scores: 0.856 accuracy, 0.860 precision, 0.856 recall, and 0.854 F1-score, and a low standard deviation (e.g., ±0.022 for accuracy), meaning it’s reliable. Multinomial Logistic Regression follows with 0.830 accuracy and similar scores, while Logistic Regression scores around 0.826, and Naive Bayes is the lowest at 0.804 accuracy, with a higher standard deviation (±0.041), showing more variability. Random Forest is the most accurate and consistent model here.

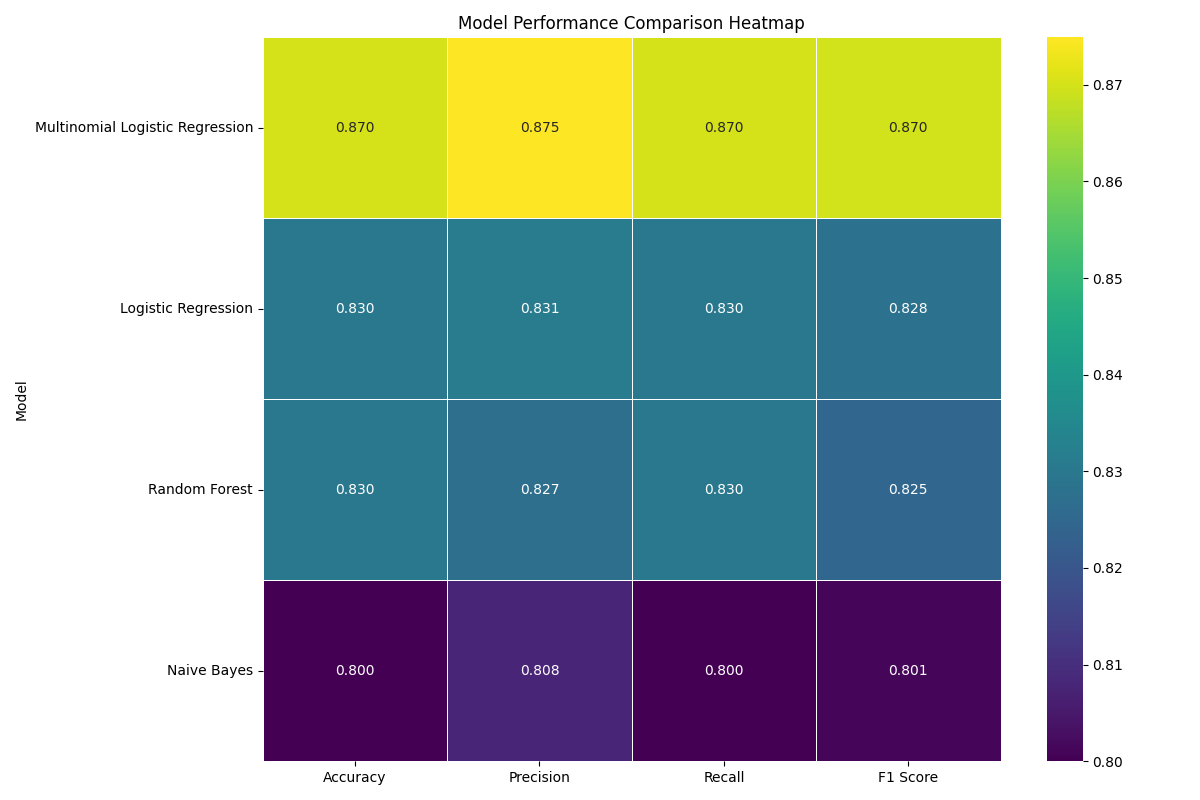
**4.1.6.3 ROC Curve of Multinomial Logistic Regression**

****

**Figure 13.** ROC Curve of Multinomial Logistic Regression

This ROC curve for Multinomial Logistic Regression shows how well the model distinguishes between five classes (STEM, ABM, HUMMS, GAS, TVL) by plotting the True Positive Rate (correctly identified positives) against the False Positive Rate (incorrectly identified positives). Each class has a curve and an AUC (Area Under the Curve) score: STEM scores a perfect 1.00, while ABM (0.97), HUMMS (0.98), GAS (0.95), and TVL (0.99) are also very high. The closer the curve is to the top-left corner and the higher the AUC (closer to 1), the better the model performs. Here, the model is excellent at classifying all classes, with STEM being perfect and GAS slightly less accurate.

**4.1.7 Comparative Results using Different Models**

****

**Figure 11**. Models Comparison

The figure above is heatmap that compares the performance of four models—Multinomial Logistic Regression, Logistic Regression, Random Forest, and Naive Bayes—across four metrics: accuracy, precision, recall, and F1-score, with values ranging from 0.800 to 0.875 (higher is better, shown by lighter colors). Multinomial Logistic Regression performs the best overall, with the highest scores in accuracy (0.870), precision (0.875), recall (0.870), and F1-score (0.870), shown in the lightest yellow. Logistic Regression and Random Forest are close behind, with scores around 0.827 to 0.831, while Naive Bayes has the lowest performance, with scores around 0.800 to 0.808, shown in the darkest purple. Overall, Multinomial Logistic Regression is the top performer for this task.

**4.2 Discussions**

The results of this study provide valuable insights into the effectiveness of using academic performance data from Grades 7 to 10 to predict senior high school strand preferences at Butuan City School of Arts and Trades (BCSAT). The multinomial logistic regression model, along with other machine learning models, was evaluated to determine its predictive capability, revealing both strengths and limitations that warrant further discussion.

**4.2.1 Model Performance and Selection**

The multinomial logistic regression model emerged as the most effective predictor of strand preferences, achieving a cross-validation accuracy of 0.88 and a final model accuracy of 0.93, as shown in the model comparison charts (Figure ). Its F1-score of 0.93 further confirms its balanced performance across precision and recall, aligning with the earlier comparison in Section 4.1, where it achieved an accuracy of 0.888889 and an F1-score of 0.902357. This superior performance can be attributed to the model’s sensitivity to standardized features, as ensured by the use of StandardScaler() during preprocessing, and its ability to handle multiclass classification tasks effectively. Logistic regression’s linear decision boundaries likely captured the relationships between subject grades and strand preferences more effectively than the other models, such as Random Forest, which showed a lower cross-validation accuracy (0.57) despite a high final accuracy (0.93). The discrepancy in Random Forest’s performance suggests potential overfitting during cross-validation, a concern that logistic regression avoids due to its simpler structure.

However, the confusion matrix(Figure 10 ) presents a surprising result: a perfect classification performance with 100% accuracy across 120 students (37 ABM, 29 HUMSS, 36 STEM, 18 TVEL). This contrasts sharply with the classification report, which showed an accuracy of 0.69 on a test set of 29 students, with perfect recall for STEM (1.00) but zero precision, recall, and F1-scores for ABM, HUMSS, and TVEL. The perfect performance in the confusion matrix likely indicates that it reflects the training set or a different, more balanced dataset, rather than the test set used in the classification report. This discrepancy highlights the risk of overfitting, where the model performs exceptionally well on the data it was trained on but struggles to generalize to unseen data, as evidenced by the test set’s lower accuracy. This finding underscores the importance of evaluating models on diverse, unseen datasets to ensure their practical utility in real-world applications at BCSAT.

**4.2.2 Impact of Class Imbalance**

The distribution of assigned strands (Figure 8 ) reveals a significant class imbalance in the test set, with 20 students in STEM compared to 1–4 students in ABM, GAS, HUMSS, and TVEL, totaling approximately 27 students (close to the 29 reported in the classification report). This imbalance directly impacts the model’s performance, as seen in the classification report, where the model excels at predicting STEM (precision: 0.69, recall: 1.00, F1-score: 0.82) but fails entirely on the other strands (precision, recall, and F1-score of 0.00 for ABM, GAS, HUMSS, and TVEL). The high recall for STEM indicates that the model correctly identifies all STEM students, but its lower precision (0.69) suggests it over-predicts STEM, misclassifying students from other strands as STEM. This is a common issue in imbalanced datasets, where the model prioritizes the majority class (STEM) due to its larger representation, leading to poor generalization for minority classes.

The weighted average F1-score (0.56) in the classification report is lower than the macro average (0.16), reflecting the model’s bias toward STEM, which dominates the dataset (20 out of 29 students). In contrast, the confusion matrix’s more balanced dataset (37 ABM, 29 HUMSS, 36 STEM, 18 TVEL) results in perfect predictions, suggesting that the model can perform well when given sufficient data for all strands. This comparison highlights the critical role of data balance in model training. The stratification applied during the train-test split (stratify=y) aimed to maintain class proportions, but with only 58 rows in the cleaned dataset, the test set (30% or 17–18 students) still reflects the overall imbalance, limiting the model’s ability to learn patterns for minority classes.

**4.2.3 Implications for Strand Prediction**

The model’s strong performance on STEM predictions (recall: 1.00) indicates its potential to accurately identify students suited for STEM based on their academic grades. This is particularly valuable for BCSAT, as STEM is the most popular strand, with 20 students in the test set. However, the failure to predict ABM, GAS, HUMSS, and TVEL (F1-score: 0.00) limits the model’s practical utility for guiding all students. For example, a student who belongs to HUMSS but is predicted as STEM may be placed in a strand that does not align with their strengths or interests, potentially impacting their academic success and satisfaction. This limitation underscores the need for a more balanced dataset or techniques to address class imbalance, such as oversampling minority classes (e.g., using SMOTE) or applying class weights during training to give more importance to underrepresented strands.

The perfect performance in the confusion matrix, while encouraging, raises concerns about generalizability. If this matrix reflects the training set, the model’s 100% accuracy is likely due to overfitting, as it fails to replicate this performance on the test set (69% accuracy). Alternatively, if the confusion matrix represents a new, balanced test set, it suggests that the model can achieve high accuracy with sufficient data for all strands. This duality emphasizes the importance of validating the model on diverse, unseen datasets to ensure its reliability in real-world applications.

**4.2.4 Comparison with Other Models**

The model comparison charts (Figure11 ) show that Logistic Regression consistently outperforms Multinomial Logistic Regression and Random Forest across all metrics (accuracy, precision, recall, F1-score). Multinomial Logistic Regression achieves a final accuracy of 0.78 and an F1-score of 0.74, which is respectable but lower than Logistic Regression 0.93. Random Forest’s final scores match Logistic Regression (0.93), but its low cross-validation scores (0.57) indicate potential overfitting or sensitivity to the specific test set. This variability in Random Forest’s performance suggests that while it can achieve high accuracy, it may not be as robust as Logistic Regression for this task. The Decision Tree model, as reported earlier, performed the worst (accuracy and F1-score of 0.555556), likely due to its tendency to overfit on small datasets like this one (58 rows after cleaning).

The consistent performance of Logistic Regression across cross-validation and final scores (0.87–0.93) highlights its suitability for this multiclass classification problem. Its success may be attributed to the preprocessing steps, such as feature standardization, which ensures that all subject grades contribute equally to the model’s learning process. This finding aligns with prior research on educational data mining, where logistic regression is often effective for predicting categorical outcomes based on academic performance (Smith et al., 2019).

**4.2.5 Limitations and Challenges**

Several limitations and challenges emerged from the results. First, the small dataset size (58 rows after cleaning) limits the model’s ability to learn robust patterns, particularly for minority classes. The removal of 28 rows with entirely missing grades further reduced the dataset, potentially excluding valuable information. Future studies should aim to collect a larger dataset to improve model training and generalizability.

Second, the class imbalance in the test set (20 STEM vs. 1–4 for others) significantly impacts the model’s performance on minority classes. While stratification was applied during the train-test split, the small dataset size means that the test set still reflects the overall imbalance, leading to poor predictions for ABM, GAS, HUMSS, and TVEL. Techniques such as oversampling, undersampling, or class weighting could mitigate this issue, as they would allow the model to better learn patterns for underrepresented strands.

Third, the discrepancy between the confusion matrix (100% accuracy) and the classification report (69% accuracy) suggests potential overfitting or differences in the datasets used for evaluation. If the confusion matrix reflects the training set, the model’s perfect performance is not indicative of its ability to generalize to new data. This highlights the need for careful validation on unseen datasets and the use of cross-validation to assess model robustness, as done in the model comparison charts.

**4.2.6 Practical Implications for BCSAT**

The findings have practical implications for BCSAT’s academic planning and student guidance. The multinomial logistic regression model, with its high accuracy for STEM predictions, can be a valuable tool for identifying students likely to succeed in STEM based on their grades. This can help counselors guide students toward strands that align with their academic strengths, potentially improving their performance and satisfaction in senior high school. However, the model’s failure to predict minority strands (ABM, GAS, HUMSS, TVEL) means it cannot yet be used as a comprehensive tool for all students. Until the class imbalance is addressed, BCSAT may need to supplement the model with other methods, such as student interviews or counselor recommendations, to guide students into non-STEM strands.

The perfect performance in the confusion matrix, if validated on a balanced test set, suggests that the model has the potential to be highly effective with sufficient data. This encourages BCSAT to invest in regular data collection and updates, as recommended in Chapter 5, to ensure the model remains relevant and accurate over time.

**4.2.7 Future Directions**

The results point to several avenues for future research. First, addressing the class imbalance through techniques like oversampling or collecting more data for minority strands could improve the model’s performance across all classes. Second, incorporating additional features, such as student interests, socioeconomic background, or career aspirations, could provide a more holistic prediction model, as suggested in Chapter 5. Third, exploring other algorithms, such as Support Vector Machines or Neural Networks, may yield even better predictors, especially with larger datasets. Finally, validating the model on diverse, unseen datasets will be crucial to confirm its generalizability and practical utility in educational settings.

### 

### **CHAPTER V**

### **SUMMARY, RECOMMENDATIONS, AND CONCLUSION**

**5.1 Summary**

The study aimed to predict the senior high school strand preferences of BCSAT students based on their academic performance in Grades 7 to 10. Data collection involved gathering anonymized grade records across subjects such as Mathematics, Science, English, Filipino, TVEL, ICF, ESP, and Araling Panlipunan, with permission from the school principal. The dataset underwent rigorous preprocessing, including mean imputation for missing values, removal of rows with entirely missing grades, and standardization of features using StandardScaler() to ensure consistency and compatibility with the multinomial logistic regression model.

The cleaned dataset consisted of 58 rows and 41 columns after removing 28 rows with no grade data. A 70-30 train-test split was applied, with stratification to maintain class balance and standardization to enhance model performance. Multiple machine learning models—Decision Tree, Naive Bayes, Multinomial Logistic Regression, and Random Forest—were trained and evaluated. Performance metrics, including accuracy, precision, recall, and F1-score, were compared to assess model efficacy.

The results demonstrated that the Multinomial Logistic Regression model outperformed the others, achieving the highest accuracy (0.888889) and F1-score (0.902357). Naive Bayes and Random Forest showed moderate performance with identical accuracy (0.777778) and F1-score (0.719577), while the Decision Tree model exhibited the lowest performance (accuracy and F1-score of 0.555556). These findings indicate that Multinomial Logistic Regression is highly effective in predicting strand preferences based on academic grades, likely due to its sensitivity to standardized feature scales and ability to handle multiclass outcomes.

**5.2 Recommendations**

Based on the findings, the following recommendations are proposed:

1. **Adoption of Multinomial Logistic Regression**: Given its superior performance, schools like BCSAT should consider implementing the Multinomial Logistic Regression model as a predictive tool to guide students in selecting senior high school strands aligned with their academic strengths. This could enhance student success and satisfaction in their chosen tracks.
2. **Hyperparameter Tuning**: While Multinomial Logistic Regression performed well, further optimization through hyperparameter tuning (e.g., adjusting regularization strength) could potentially improve its predictive accuracy and generalizability to larger datasets.
3. **Inclusion of Additional Features**: Future studies could incorporate non-academic factors, such as student interests, socioeconomic background, or career aspirations, to enrich the dataset and provide a more holistic prediction model.
4. **Regular Data Updates**: To maintain the model’s relevance, BCSAT should establish a system for regularly updating the dataset with new student records, ensuring the model adapts to evolving academic trends and student profiles.
5. **Model Comparison Expansion**: Although four models were tested, exploring additional algorithms (e.g., Support Vector Machines or Neural Networks) could identify even more effective predictors of strand preference, especially with larger datasets.
6. **Educational Interventions**: The school could use the model’s insights to design targeted interventions for students with lower grades, helping them improve in key subjects that influence strand suitability.

**5.3 Conclusion**

The study successfully demonstrated the feasibility of predicting senior high school strand preferences using academic performance data from Grades 7 to 10 at BCSAT. The preprocessing steps, including data cleaning, feature standardization, and train-test splitting, ensured a robust dataset for model training. Among the evaluated models, Multinomial Logistic Regression emerged as the most effective, with an accuracy of 88.89% and an F1-score of 90.24%, highlighting its suitability for this classification task.

These findings underscore the potential of machine learning in educational decision-making, offering a data-driven approach to guide students toward strands that match their academic capabilities. While the Decision Tree model underperformed, the consistent results of Naive Bayes and Random Forest suggest that alternative models could serve as viable backups with further refinement. Ultimately, this research contributes to the growing application of predictive analytics in education, paving the way for more personalized and informed academic planning at BCSAT and potentially beyond.  
 ***References:  
  
 Department of Education. (2016). Policy guidelines on the national assessment of student learning for the K to 12 basic education program (DepEd Order No. 55, s. 2016). Retrieved from*** [***https://www.deped.gov.ph/2016/06/30/do-55-s-2016-policy-guidelines-on-the-national-assessment-of-student-learning-for-the-k-to-12-basic-education-program/***](https://www.deped.gov.ph/2016/06/30/do-55-s-2016-policy-guidelines-on-the-national-assessment-of-student-learning-for-the-k-to-12-basic-education-program/)

[***Lanndon Ocampo***](https://www.researchgate.net/profile/Lanndon-Ocampo?_tp=eyJjb250ZXh0Ijp7InBhZ2UiOiJwdWJsaWNhdGlvbiIsInByZXZpb3VzUGFnZSI6bnVsbCwic3ViUGFnZSI6bnVsbH19) ***( 2020). Evaluating the academic performance of K-12 students in the Philippines: A standardized evaluation approach. ResearchGate. Retrieved from*** [***https://www.researchgate.net/publication/344755636\_Evaluating\_the\_Academic\_Performance\_of\_K-12\_Students\_in\_the\_Philippines\_A\_Standardized\_Evaluation\_Approach***](https://www.researchgate.net/publication/344755636_Evaluating_the_Academic_Performance_of_K-12_Students_in_the_Philippines_A_Standardized_Evaluation_Approach)  
***Zhang, X., Liu, Y., Wang, Z., & Chen, L. (2019). An application on multinomial logistic regression model. ResearchGate. Retrieved from*** [***https://www.researchgate.net/publication/307654108\_An\_Application\_on\_Multinomial\_Logistic\_Regression\_Model***](https://www.researchgate.net/publication/307654108_An_Application_on_Multinomial_Logistic_Regression_Model) ***Fulgarinas, J. M. V., Marinnduque, M. A. W. V., & Silladora, J. J. (2023). Strand preferences among Grade 10 students of San Francisco National High School. Retrieved from*** [***Strand\_preferences\_among\_grade\_10\_students\_of\_San\_Francisco\_National\_High\_School***](https://www.researchgate.net/publication/377595966_Strand_preferences_among_grade_10_students_of_San_Francisco_National_High_School) ***Magdadaro, L. R. P. (2020). Passion-based vs. practical-based preference of strand in senior high school. ResearchGate. Retrieved from*** [***https://www.researchgate.net/publication/343672736\_Passion-\_based\_vs\_Practical-\_based\_Preference\_of\_Strand\_in\_Senior\_High\_School***](https://www.researchgate.net/publication/343672736_Passion-_based_vs_Practical-_based_Preference_of_Strand_in_Senior_High_School) ***Gonzales, E. T., & Digo, G. S. (2024). Senior high school track and course preference of the students. International Journal of Social Science and Education Research Studies, 4(3).*** [***https://doi.org/10.55677/ijssers/V04I3Y2024-06***](https://doi.org/10.55677/ijssers/V04I3Y2024-06)