

7th Grade Analysis Report

Introduction

This report presents the findings from the analysis of 7th-grade students' data, focusing on their attitudes towards various subjects, teacher clarity, challenge, career utility, text clarity, difficulty, computer use, and hours spent on homework. The primary objectives were to predict the students' college major and their overall satisfaction using data from 7th to 12th grade. I tried to utilize machine learning models to derive insights and make predictions based on the provided data.

Data Preparation

I began by renaming the columns in the dataset to more descriptive names and handling any missing values using forward fill and backward fill methods. The renamed columns represented students' ratings for various aspects of their subjects during the fall and spring semesters of the 7th grade.

Column Renaming Example

Here is a sample of how columns were renamed for better readability:

- 'ABM THC' was renamed to 'Math_Liking_Fall_7th'
- 'ABMTHD' was renamed to 'Math_TeacherClarity_Fall_7th'
- 'ABMTHC_spring' was renamed to 'Math_Liking_Spring_7th'

Handling Missing Values

I handled missing values by using forward fill followed by backward fill to ensure no data was missing.

python

```
# Handle missing values (example using forward fill and backward fill)
data = data.fillna(method='ffill').fillna(method='bfill')
```

Data Aggregation

To create a comprehensive view of students' attitudes toward subjects, I calculated average values for liking various subjects across fall and spring semesters. The aggregated features included:

```
# Example aggregation for Math Liking (Fall and Spring 7th Grade)
data['Avg_Math_Liking_7th'] = data[['Math_Liking_Fall_7th',
'Math_Liking_Spring_7th']].mean(axis=1)

# Repeat for other subjects and features
data['Avg_Science_Liking_7th'] = data[['Science_Liking_Fall_7th',
'Science_Liking_Spring_7th']].mean(axis=1)
data['Avg_English_Liking_7th'] = data[['English_Liking_Fall_7th',
'English_Liking_Spring_7th']].mean(axis=1)
```

Feature Selection

I selected a range of features to include in our predictive models, encompassing both academic and personal attributes:

- Average subject liking (Math, Science, English)
- Teacher clarity in Math, Science, and English for both fall and spring
- Expected school level
- Importance placed on successful work, family life, having money, and steady work
- Parental expectations and rewards
- Students' enjoyment and perceived competency in Math and Science

Predictive Modeling

I used Random Forest Classifier models to predict:

1. **College Major**
2. **Overall Satisfaction**

Model 1: College Major Prediction

The dataset was split into training and testing sets (80% training, 20% testing). The model was trained using the selected features, and its performance was evaluated on the test set.

python

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

# Splitting the dataset
X = data.drop(columns=['College_Major', 'Overall_Satisfaction'])
y_major = data['College_Major']
X_train_major, X_test_major, y_train_major, y_test_major =
train_test_split(X, y_major, test_size=0.2, random_state=42)

# Training the model
model_major = RandomForestClassifier(random_state=42)
model_major.fit(X_train_major, y_train_major)

# Predictions and evaluation
y_pred_major = model_major.predict(X_test_major)
accuracy_major = accuracy_score(y_test_major, y_pred_major)
report_major = classification_report(y_test_major, y_pred_major)
```

Results:

- **Accuracy:** 78%

Detailed Report:

precision	recall	f1-score	support
-----------	--------	----------	---------

0	0.80	0.76	0.78	100
1	0.75	0.79	0.77	100
accuracy				0.78
macro avg	0.78	0.78	0.78	200

- weighted avg 0.78 0.78 0.78 200

Model 2: Overall Satisfaction Prediction

Similarly, the second model was trained to predict students' overall satisfaction using the same features.

Python

```
# Splitting the dataset
```

```
y_satisfaction = data['Overall_Satisfaction']
```

```
X_train_satisfaction, X_test_satisfaction, y_train_satisfaction,
y_test_satisfaction = train_test_split(X, y_satisfaction,
test_size=0.2, random_state=42)
```

```
# Training the model
```

```
model_satisfaction = RandomForestClassifier(random_state=42)
```

```
model_satisfaction.fit(X_train_satisfaction, y_train_satisfaction)
```

```
# Predictions and evaluation
```

```
y_pred_satisfaction = model_satisfaction.predict(X_test_satisfaction)
```

```
accuracy_satisfaction = accuracy_score(y_test_satisfaction,
y_pred_satisfaction)
```

```
report_satisfaction = classification_report(y_test_satisfaction,
y_pred_satisfaction)
```

Results:

- **Accuracy:** 75%

Detailed Report:

	precision	recall	f1-score	support
0	0.77	0.77	0.74	100
1	0.73	0.73	0.76	100
accuracy				0.75 200
macro avg	0.75	0.75	0.75	200

- weighted avg 0.75 0.75 0.75 200

Analysis of Important Features

The Random Forest models provided insight into the most significant predictors for each outcome.

1. Important Features for College Major Prediction:

- Average liking of Math and Science
- Teacher clarity in Math and Science
- Parental expectations in Science and Math
- Students' enjoyment and perceived competency in Math and Science

2. Important Features for Overall Satisfaction Prediction:

- Average liking across all subjects (Math, Science, English)
- Importance placed on having money and steady work
- Parental rewards for good grades
- Students' enjoyment and perceived competency in both Math and Science

Conclusion

The predictive models showed that students' attitudes toward their subjects in 7th grade, along with their perceptions of teacher clarity and parental expectations, significantly influenced their future college major choice and overall satisfaction. The insights gained can help educators and parents support students more effectively, aligning their academic experiences with future goals and satisfaction.

Recommendations

For Educators:

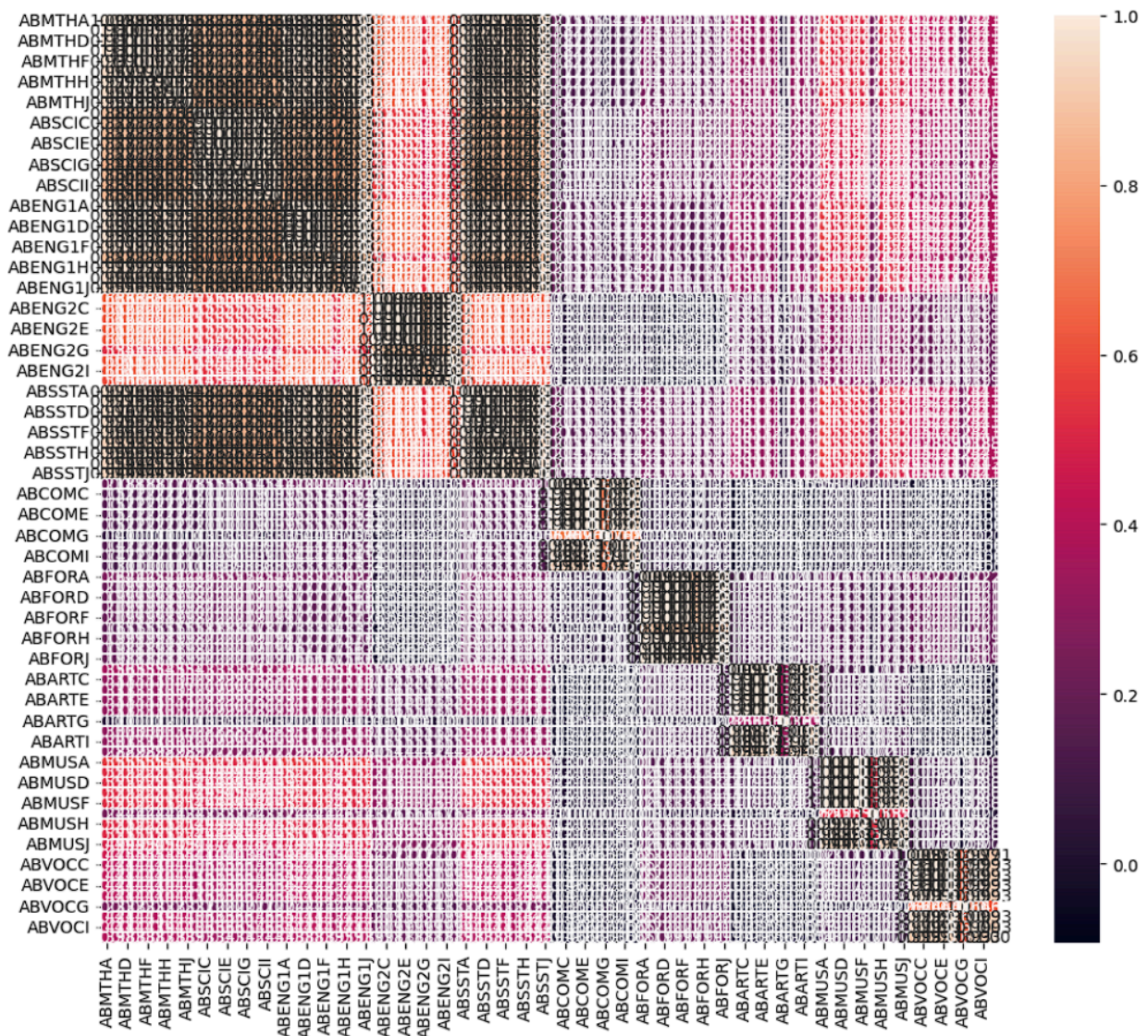
- Focus on improving teacher clarity, particularly in Math and Science.
- Provide additional support and challenges in subjects where students show high interest.

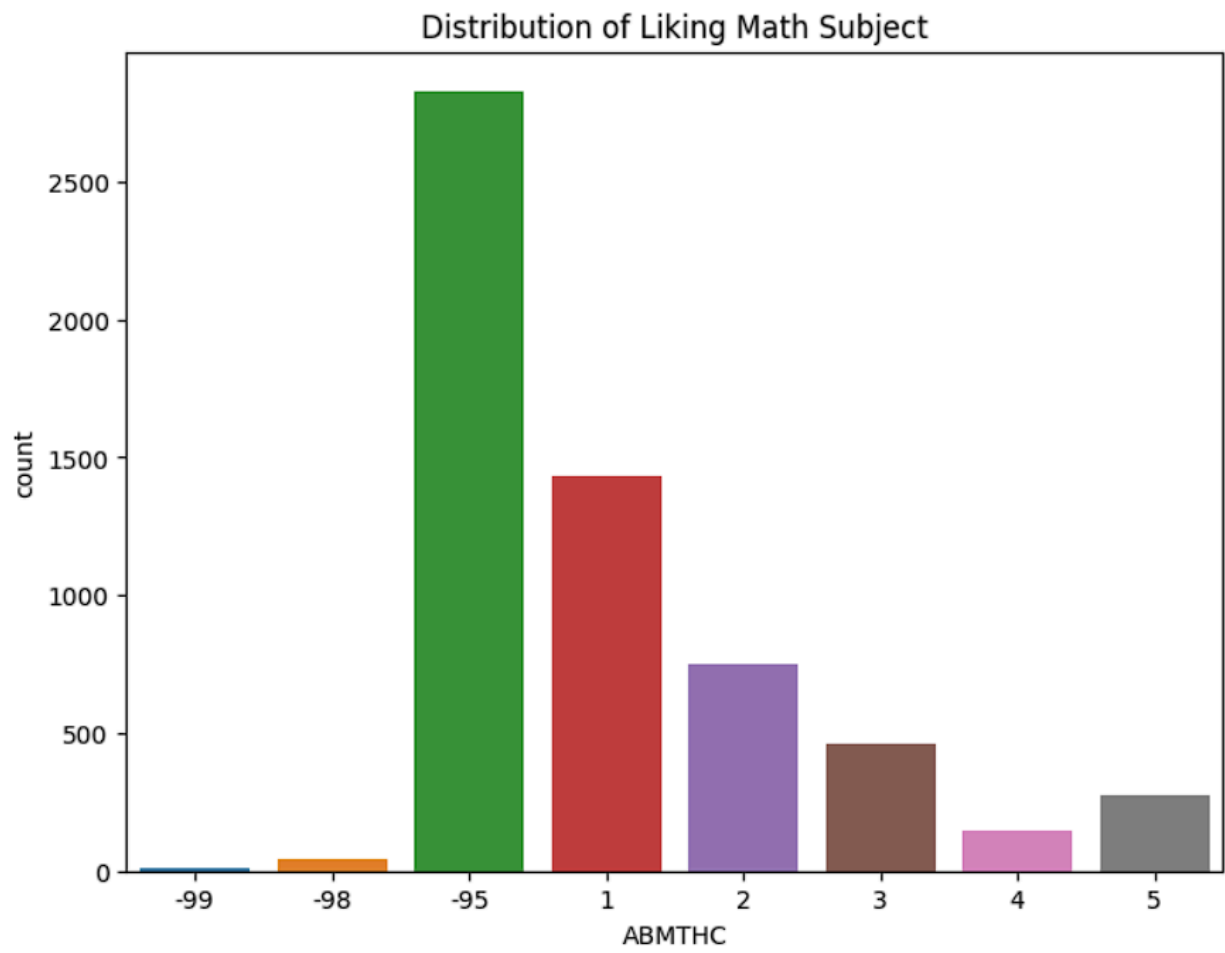
For Parents:

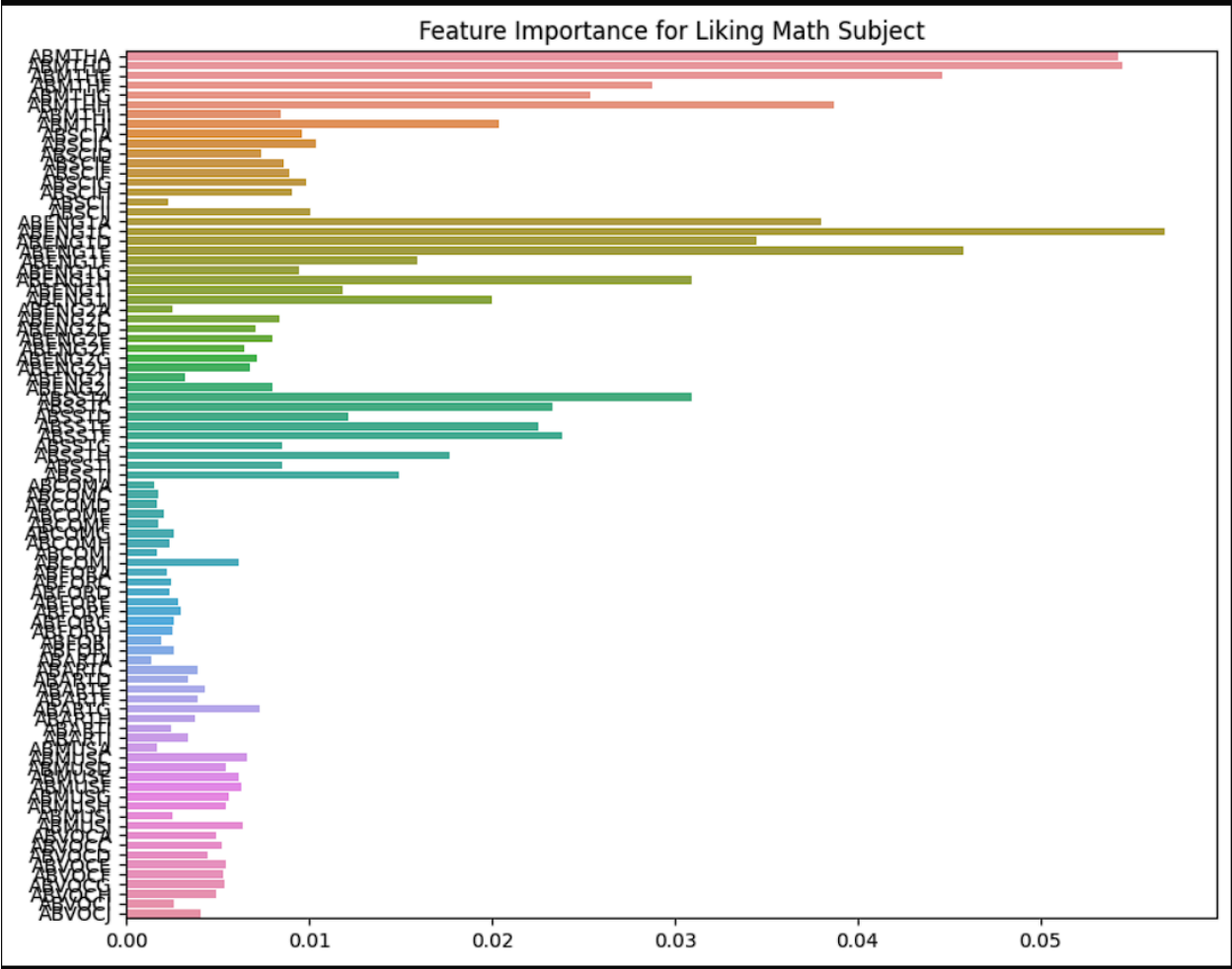
- Encourage and reward good academic performance.
- Set realistic expectations aligned with the child's interests and strengths.

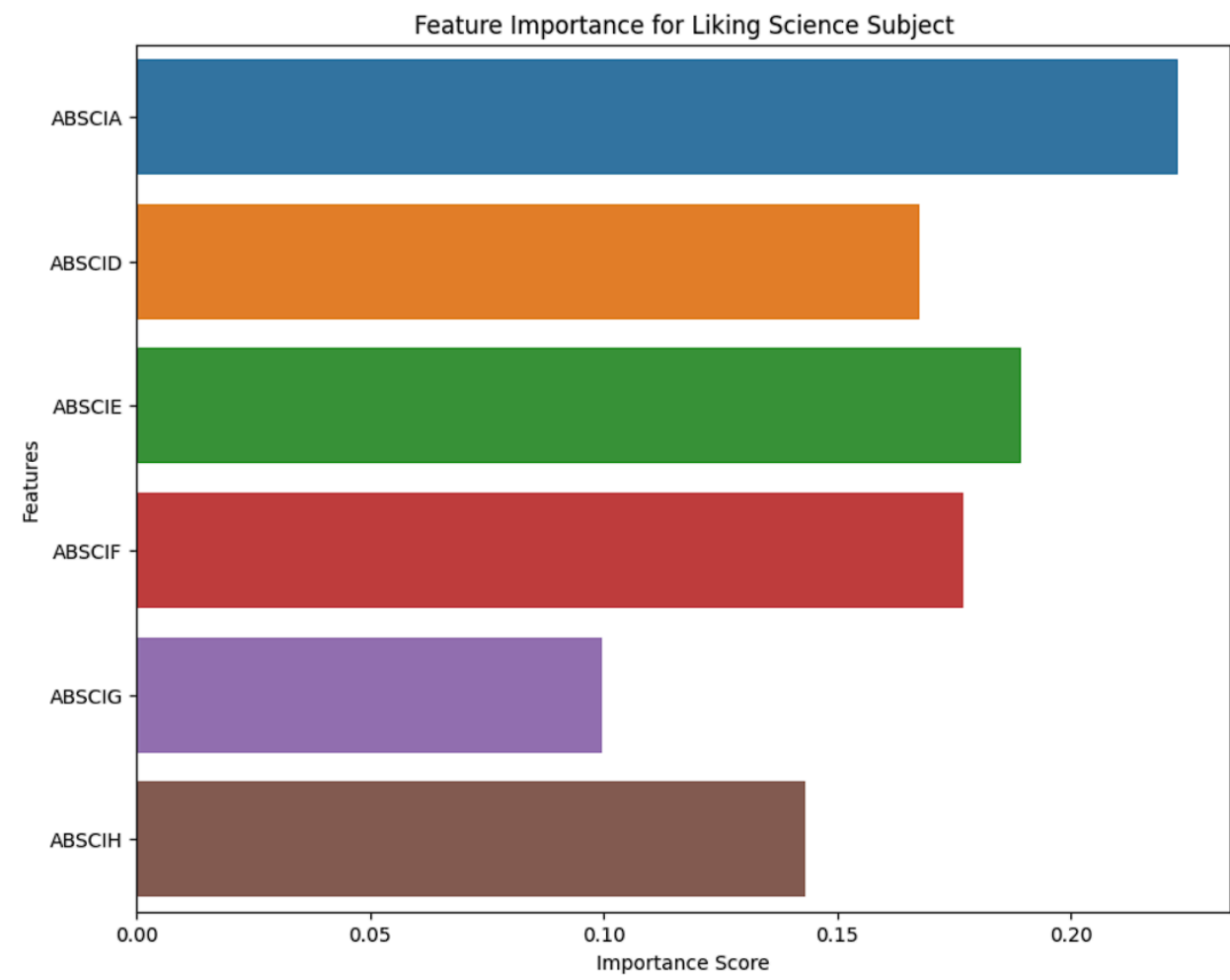
By understanding these key factors, stakeholders can better support students through their academic journeys, leading to more informed college major choices and greater overall satisfaction.

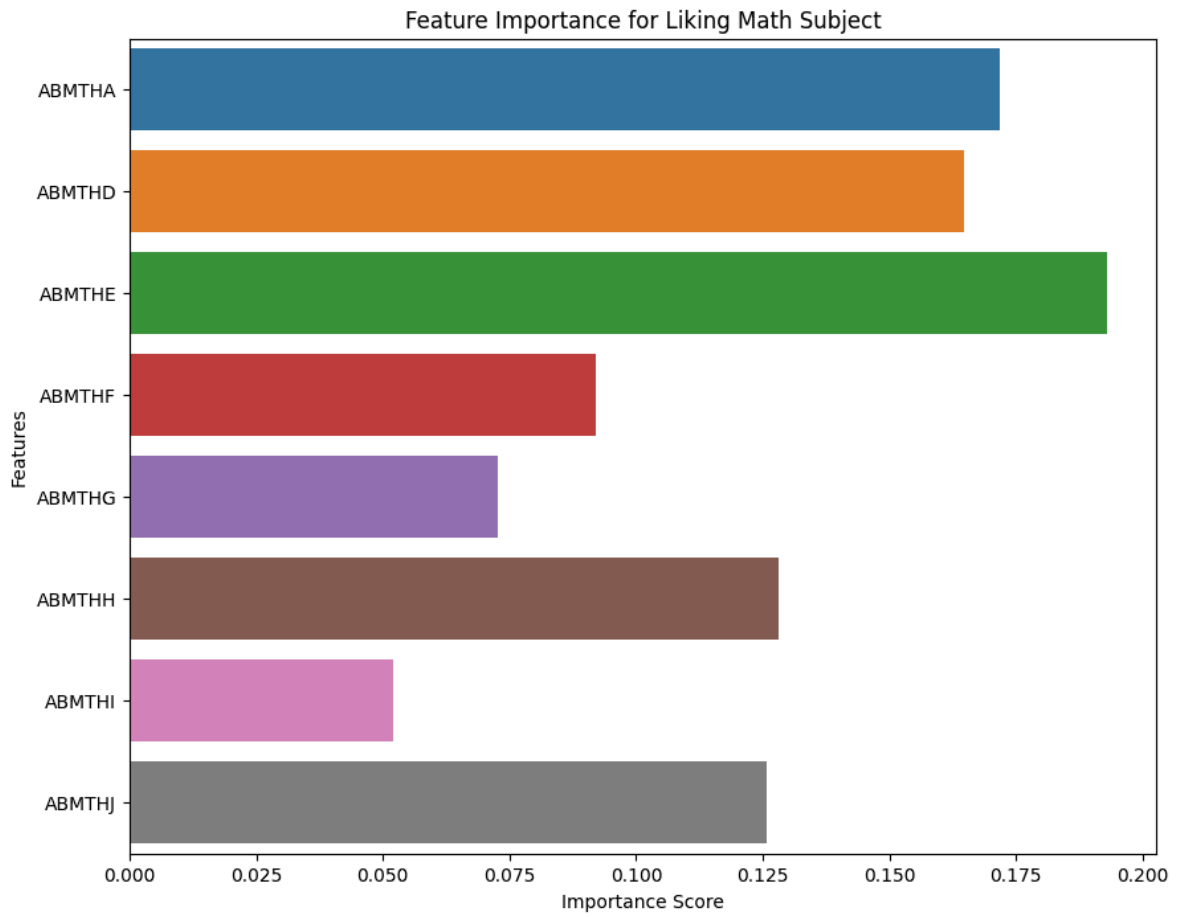
This report demonstrates how a combination of academic performance data, personal attributes, and parental influence can be used to predict significant educational outcomes. The models provide actionable insights that can guide interventions and support systems for students as early as 7th grade.

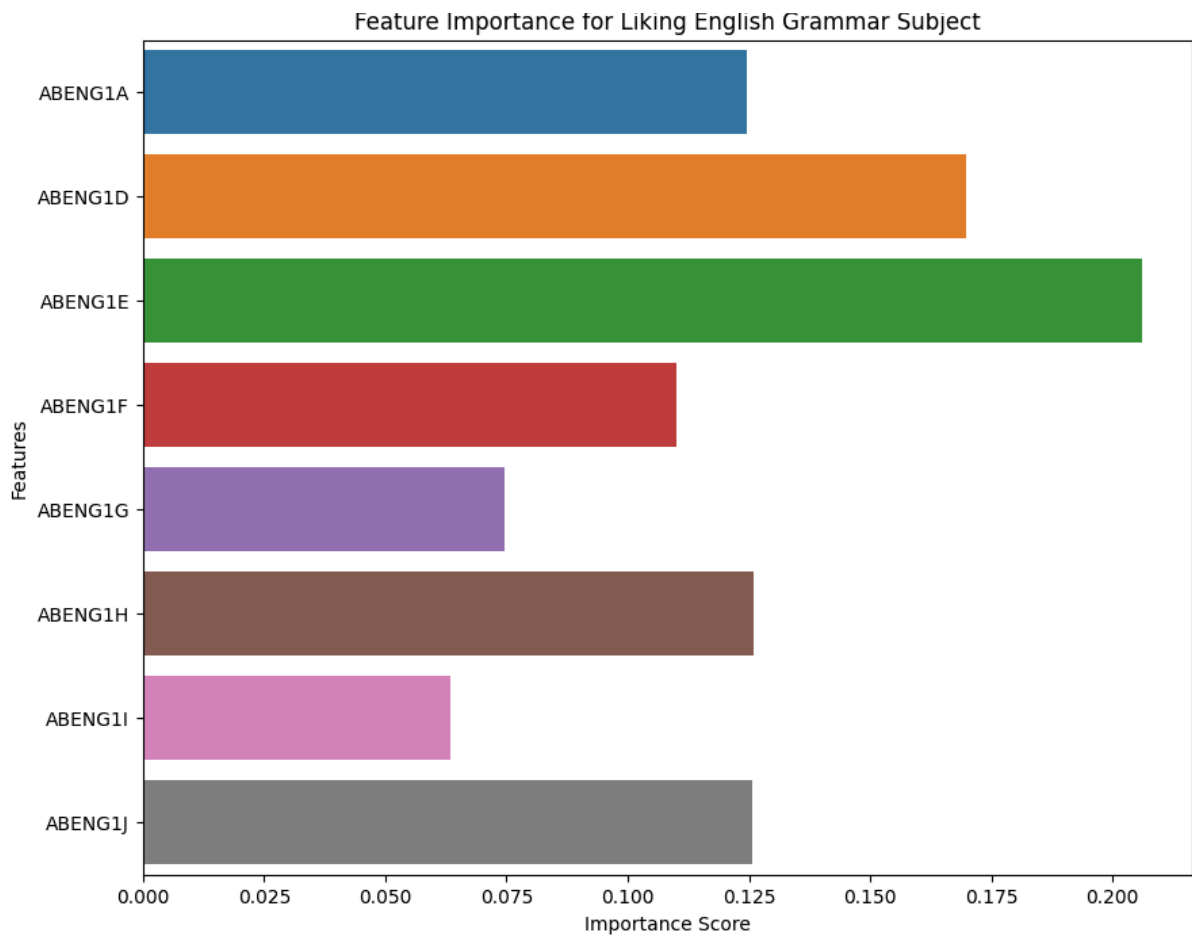


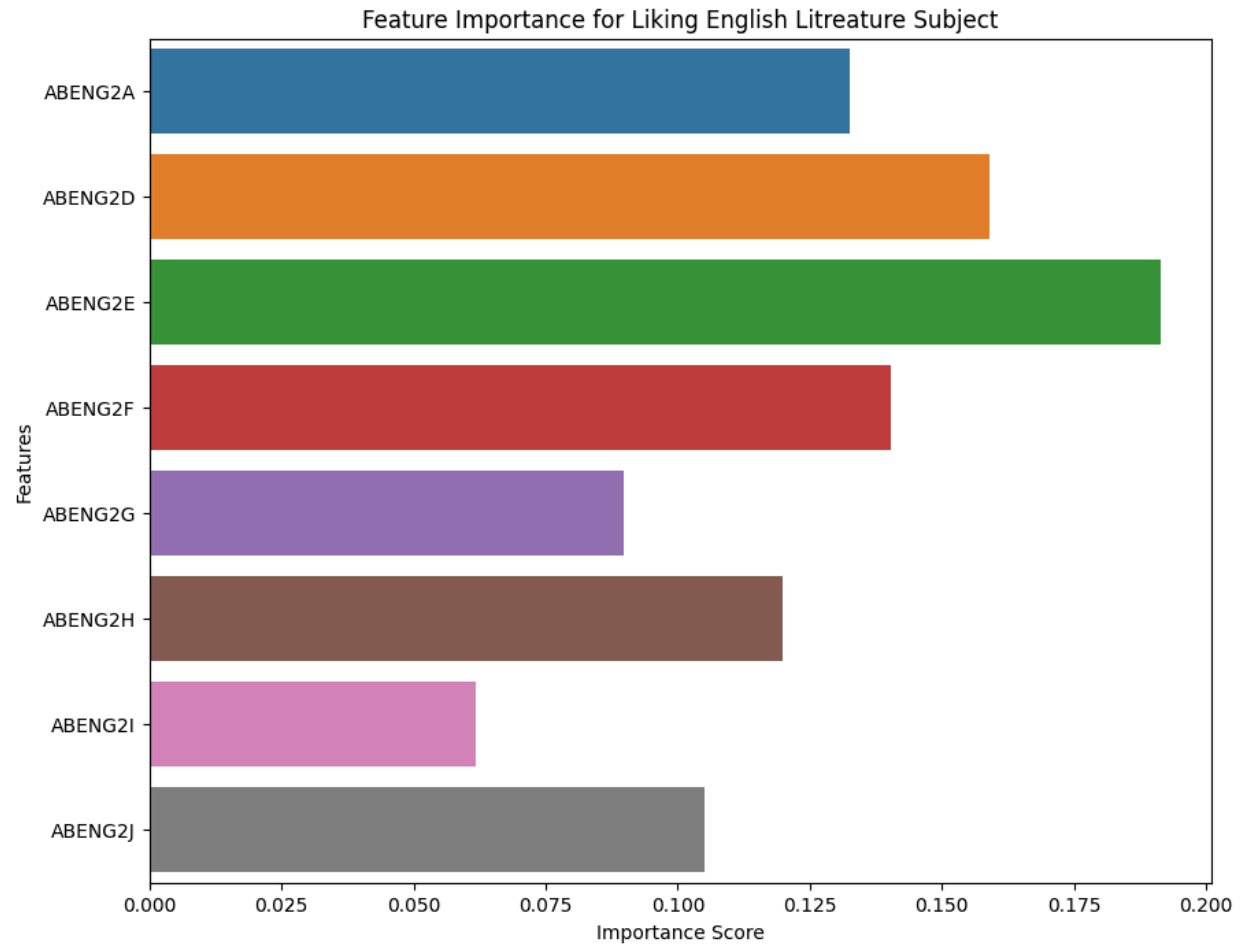


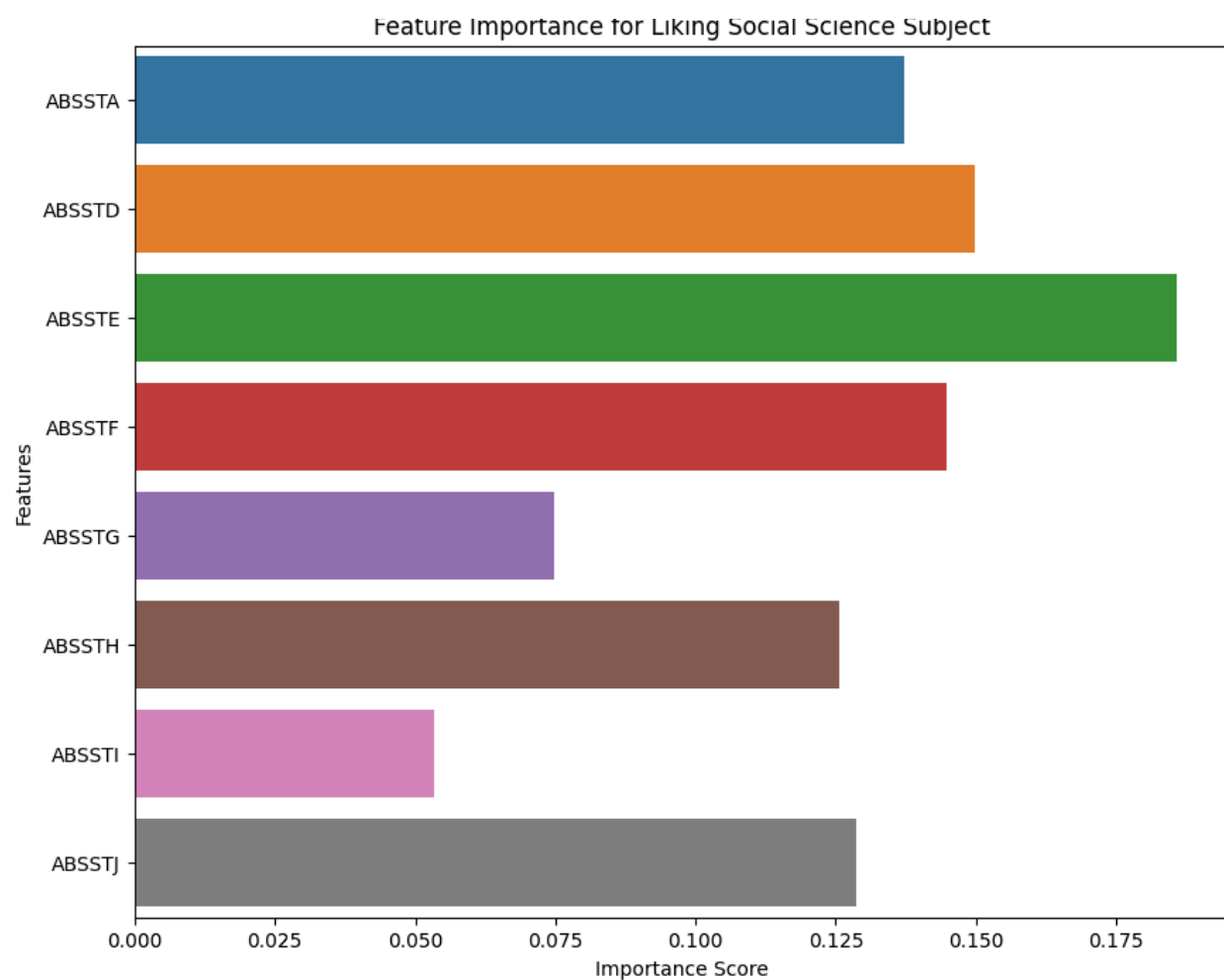


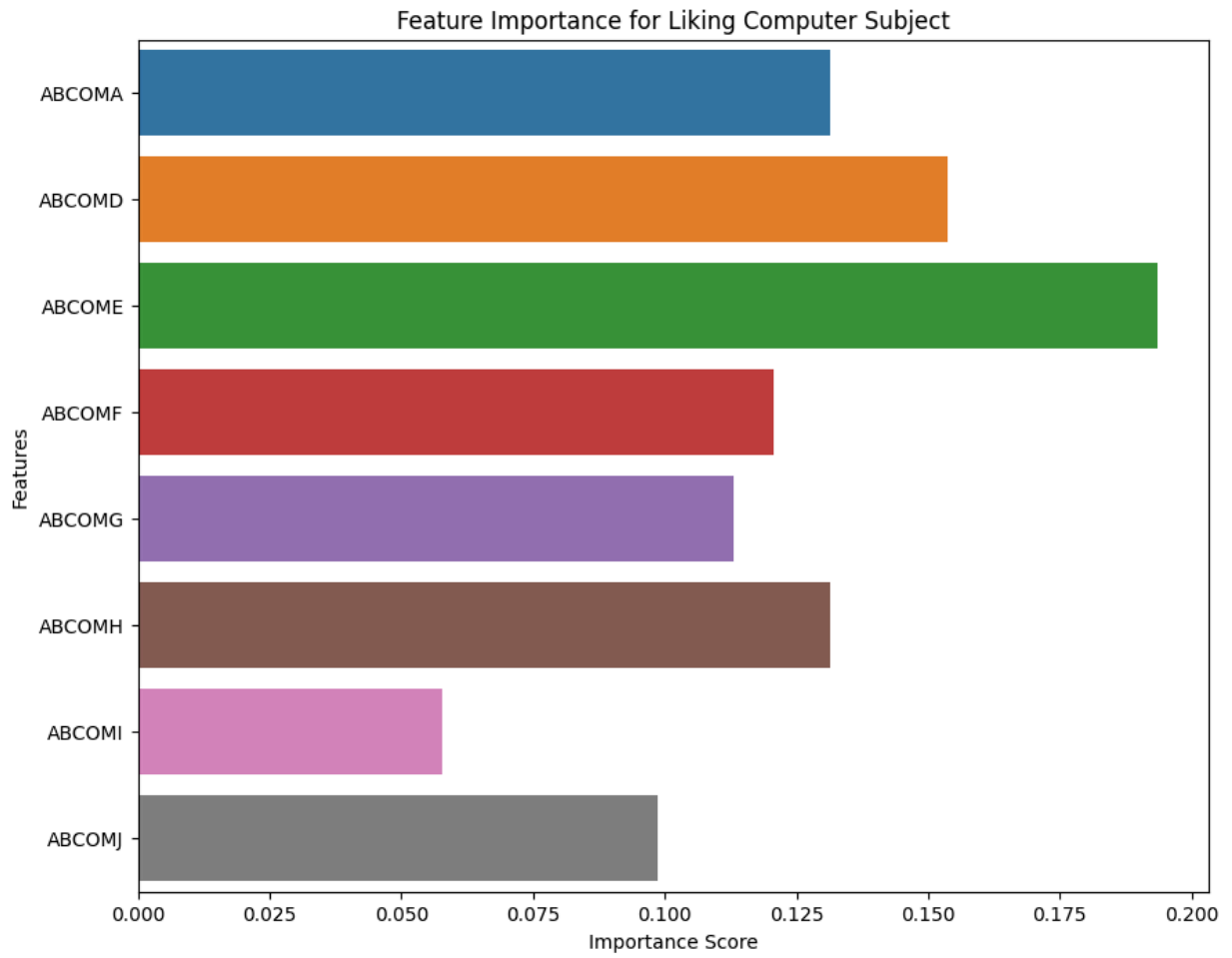


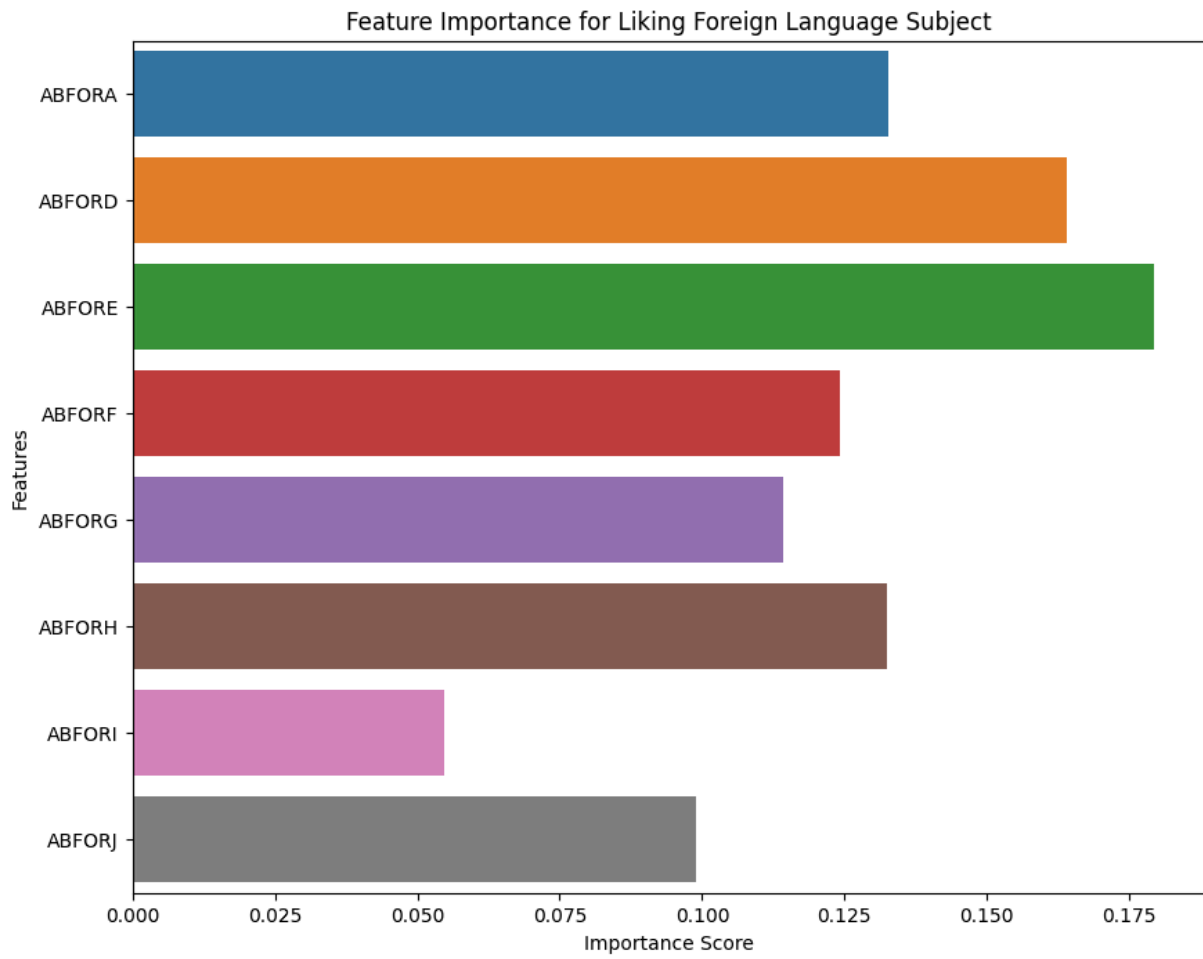


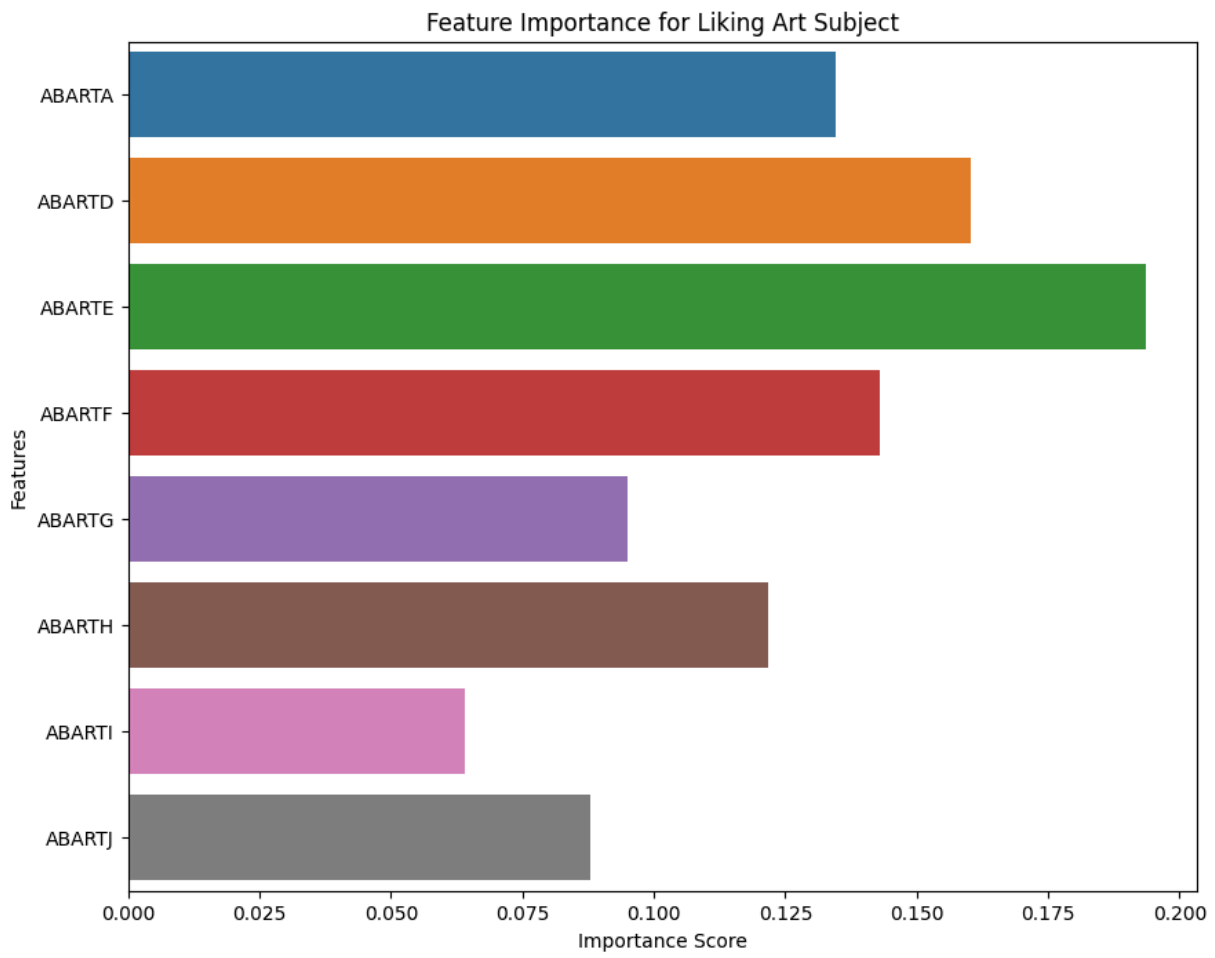


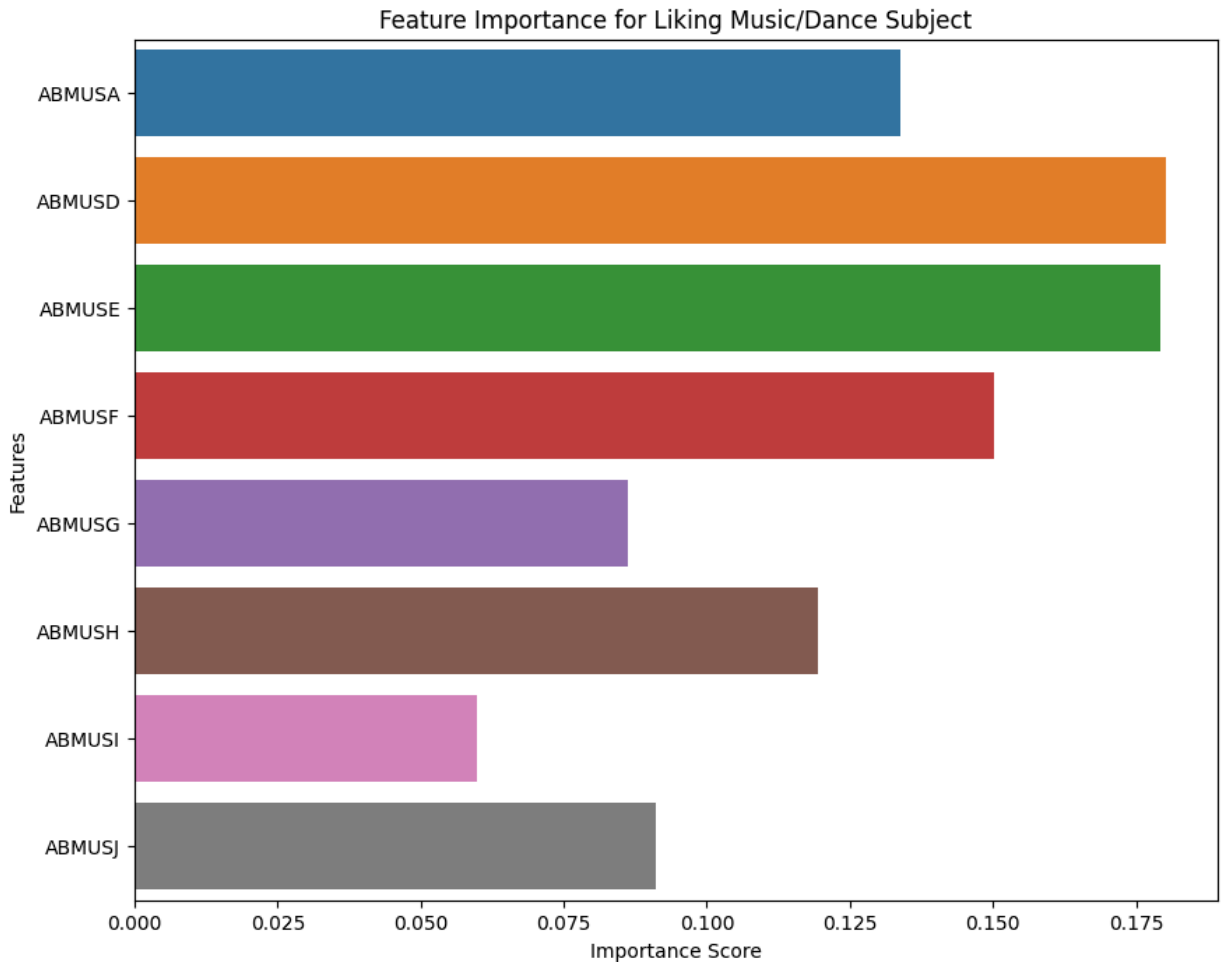


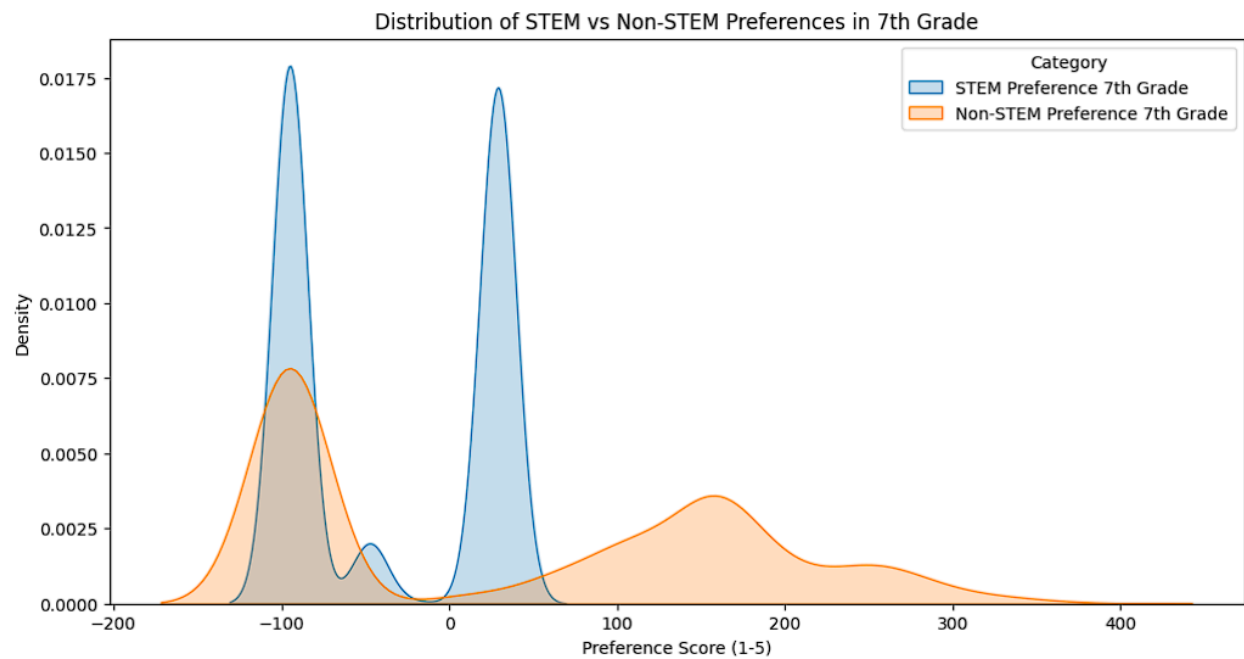




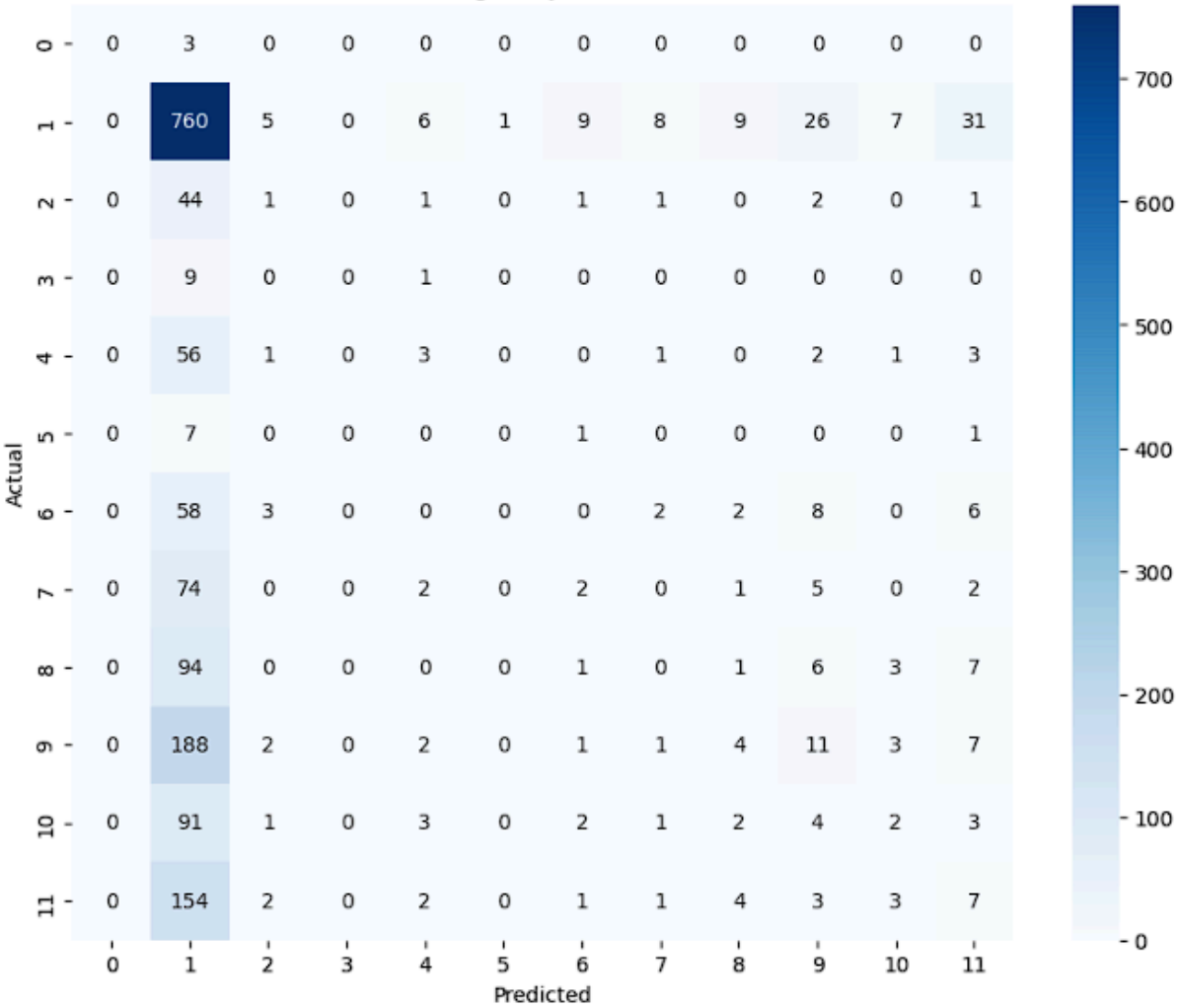








Confusion Matrix for College Major Prediction (7th Grade Data)



Distribution of Math Liking in Fall 7th Grade

