AI in Education: Navigating Fairness and Inclusivity Challenges

STUDENT NAME:	Mandara Rangaswamy
STUDENT ID:	69786092
MASTER PROGRAM:	Msc Data Science
DATE OF SUBMISSION:	02-02-2024

Content

ABSTRACT

This study investigates how learners' performance is affected by AI-driven personalization. The study shows a beneficial relationship between enhanced academic accomplishment, engagement, and satisfaction with individualized AI-based adaptive learning through quantitative and qualitative analysis. The results demonstrate how AI-driven customisation can improve student achievement and change the way that education is delivered [RAHAF ALAMRI ,2021].

It will also unavoidably have an effect on the current social structure and give rise to moral questions. People have experienced a considerable deal of distress due to ethical concerns brought about by AI systems, such as security hazards, discrimination, privacy leakage, and unemployment.

As a result, the study of ethical dilemmas in AI, or AI ethics, has grown in importance as a topic of academic research as well as a topic of general concern for people, organizations, nations, and society.[Changwu Huang,2023]

KEYWORDS : Learners' performance, Education technology, Artificial intelligence, learning outcomes, Learning Analytics, Engagements, Personalized Learning

INTRODUCTION

This research tries to investigate and mitigate issues of inclusivity and fairness related to automated grading systems, personalized learning, and student performance prediction. It also addresses the ethical implications of AI in education. In order to achieve equal educational outcomes, it is crucial that the study attempt to comprehend any potential biases and inequities that might occur.

For many parties, including students, instructors, and academic institutions, predicting students' performance in a particular course or a whole program is a crucial undertaking. Applications for predicting student performance have shown promise in identifying at-risk students and predicting dropout rates. In order to enhance the educational experience for pupils, it is also utilized to create personalized recommendation systems and early warning systems.[RAHAF ALAMRI, 2021]

Returning to the education sector, explainable models are highly valuable to a wide range of stakeholders. As an example, let us look at the task of predicting student performance. An explainable model will be helpful to at least four separate stakeholders in this work, and the advantages for each group varied.

PROBLEM STATEMENT:

Artificial intelligence (AI)-driven personalized learning platforms have emerged as a result of the quick development and integration of AI in education, offering customized learning experiences for each student. While excitement about the potential of AI-driven personalization to revolutionize education and boost students' performance is growing, more empirical data and thorough research are still required to fully comprehend how these technologies are actually affecting students' academic performance, levels of engagement, and general academic achievement.[Amit Das, 2023]

This study intends to shed light on the possible advantages and difficulties of adopting AI technology in personalised learning environments by examining these research issues and providing significant insights into the effect of AI-driven personalization on learners' performance. The results will provide guidance to educational institutions, legislators, and teachers on how to best use AI-driven personalization to improve students' academic performance and promote a more student-centered educational philosophy.

Additionally, this research will contribute to our understanding of the ethical implications and issues that need to be taken into account when putting AI-driven personalized learning systems into practice in order to provide fair and efficient learning opportunities for all students.

Research Question/Hypothesis:

The investigation is guided by well-defined research questions and hypotheses. These could involve questions regarding the efficiency of AI in tailored education, the precision of performance prediction models, and the impartiality of computerized grading schemes. Expected relationships between variables are outlined in hypotheses.

OBJECTIVE:

- Examine Bias in AI Algorithms
- Enhance Fairness and Inclusivity
- Evaluate Ethical Implementation

1. To offer ideas for putting AI-driven personalized learning into practice that are supported by evidence: With regard to the successful integration of AI-driven personalized learning, this purpose seeks to provide educators, institutions, and policymakers with actionable recommendations. The study aims to determine the most effective methods and techniques for enhancing students' performance using individualized learning approaches.[Amit Das, 2023]
2. To ascertain how the levels of engagement of learners are affected by AI-driven personalization: This goal is to comprehend how students' involvement in the learning process is affected by AI-driven personalized learning. It will investigate whether more active participation, motivation, and curiosity result from personalized learning and whether this increases learner engagement.
3. To list possible moral issues and difficulties with AI-powered individualized learning: Examining the ethical ramifications of AI-powered personalized learning platforms is the goal of this purpose. The study will look into problems with algorithmic bias, data privacy, and the appropriate application of AI in educational contexts.[Amit Das, 2023]
   Numerous reasons, including the programming codes, the input data, incorrect operation, and other aspects, could result in the unwanted outcome. This results in what is known as "the problem of many hands" [Changwu Huang,2023]. Accountability, then, is an ethical problem pertaining to the human aspects of AI design, implementation, use, and deployment.[Changwu Huang,2023]

LITERATURE REVIEW

Artificial Intelligence (AI) in personalized learning systems allows for the development of individualized paths, adaptable content, and real-time feedback mechanisms, which are revolutionizing the field of education. We compiled recent systematic studies that address two primary subjects—student performance prediction models and explainable artificial intelligence—in order to be consistent with previous work.[RAHAF ALAMRI ,2021]

Artificial Intelligence (AI) in personalized learning systems allows for the development of individualized paths, adaptable content, and real-time feedback mechanisms, which are revolutionizing the field of education. Because of their uniformity and efficiency, AI-powered automated grading systems have become more popular in educational environments.

Prior research (Baepler et al., 2014) emphasizes how critical it is to comprehend customized learning's historical context in order to appreciate its evolution. Although there have been tailored training attempts since the 19th century, the broad adoption of personalized learning models has been made easier by recent technological breakthroughs.

RESEARCH DESIGN AND METHODOLOGY

RESEARCH DESIGN AND METHODOLOGY

The study employs a quantitative methodology, including statistical analysis and machine learning techniques to examine the influence of demographic variables on student performance within the context of Data Science education. The purpose of the study is to evaluate the inclusivity and fairness of personalized learning predictions and automated grading systems. The process entails a thorough analysis of student data, taking into account a number of variables like geography, age, financial situation, and status as an international student. In order to provide a more nuanced understanding of potential biases and inequities in educational results, the study combines machine learning model training with exploratory data analysis as part of a mixed-methods research methodology.

POPULATION AND STUDY SAMPLE:

The population under consideration comprises Data Science students in the university. The study sample is drawn from this population, encompassing a diverse group of students from different locations (Berlin, London, New York, Sydney, Tokyo, Toronto, Melbourne, Los Angeles, Paris). The sample includes students of varying ages, financial statuses (upper class, middle class, lower class), and international student status. An extensive examination of the inclusion and fairness of AI-driven educational systems is made possible by the sample's diversity, which guarantees a representative portrayal of the student body.

SAMPLE SIZE AND SELECTION OF SAMPLE:

In order to achieve statistical significance, the sample size is chosen depending on the diversity of the student population. To ensure unbiased representation, students from each demographic category are chosen through random selection. To ensure that students from various places, age groups, financial situations, and statuses as international students are fairly included, stratified sampling is used. By using this method, it is guaranteed that the sample accurately represents the student body as a whole.

SOURCES OF DATA:

Source: The dataset consists of information on Data Science students, including demographic details and marks in various subjects.
Collection Method: Data will be collected from university records and databases, ensuring the anonymity and privacy of students.
Period: The data spans the last academic years to capture a diverse range of student experiences.
Dependent Variable: Student performance, measured by marks in SQL, Excel, Python, Power BI, and English.
Independent Variables: Location, age, address (urban/rural), financial status, and international student status.
Data Preprocessing:
Missing values, outliers, and inconsistencies will be addressed.
Sensitive information will be anonymized to comply with privacy regulations.
Categorical variables will be encoded appropriately, and numerical features will be scaled if necessary.

COLLECTION OF DATA:

Data collection involves extracting relevant information from academic records and databases, ensuring the anonymity and confidentiality of student information. Structured surveys may also be employed to gather additional insights into students' learning preferences and experiences. The data collection process prioritizes transparency and informed consent, providing students with clear explanations of how their data will be used.

DATA ANALYSIS STRATEGIES:

A complete strategy is provided to evaluate and mitigate issues of justice and inclusion in the context of automated grading systems and student success prediction for students studying data science. In order to comprehend the distribution of features like location, age, financial status, and marks in SQL, Excel, Python, Power BI, and English, the first step entails performing an exploratory data analysis (EDA). Finding any potential biases and trends in the dataset is the goal of this research. Next, we look at demographic representation, specifically how students are distributed among various places, age groups, and financial status categories. The dataset is examined closely for biases, and fairness analyses are performed to compare the distribution of scores among different demographic groups in order to assess grading systems.

Correlation analysis are utilized to comprehend the associations among various subjects and demographic characteristics. Metrics for inclusivity, such as representation parity and equalized odds, are defined to measure diversity and justice. In order to address potential biases, fairness-aware training approaches are incorporated into machine learning models for predicting student achievement. To ensure demographic representativity in both the training and testing sets, the dataset is divided. Models are tested on the testing set, and results are measured across different subgroups using fairness criteria.

If biases are found, mitigation procedures are put into place to guarantee fair results. These strategies could include resampling, reweighting, or modifying decision thresholds. The method takes ethical issues into account, discussing the ramifications of utilizing particular characteristics, such financial position, to predict student achievement. There is extensive documentation for every step of the process, including model training, data preprocessing, and fairness assessments. A thorough report is written outlining the conclusions, difficulties, and actions made to address concerns of inclusivity and justice. Furthermore, a framework for ongoing evaluation and development is created in order to adjust models to modifications in the student body and educational environment over time.

EXPECTED RESEARCH FINDINGS

After analyzing the study results, the machine learning model performed admirably overall, with accuracy scores of 97% and precision scores varying from 96% to 100% depending on the class. The model's effectiveness is further supported by the confusion matrix and classification report, which show strong precision and recall values for every class. This shows that the model can accurately predict students' performance in a variety of Data Science-related performance criteria.

To analyze the variability of features between sites, one-way ANOVA tests were performed for every subject mark. The p-values that were acquired indicate that there isn't a statistically significant variation in the average scores of students from various regions. This suggests that there is little variation in academic success, as measured by subject marks, depending on one's location.

There aren't many strong correlations found in the correlation matrix evaluating the relationships between various subject marks, which suggests some degree of subject independence. This is in line with what is expected in a multifaceted and diverse subject such as data science, where expertise in one area may not always substantially correspond with another.

However, a notable finding is the exceptionally high disparate impact value of 110000000000.0, signaling a potential imbalance in the impact of the model across different groups. This raises concerns about fairness and equity in the model's predictions. Addressing this issue is paramount, and further investigation into the demographic distribution of the dataset is recommended. Analyzing the model's impact on various student groups will help identify and rectify any underlying biases or disparities.

These results highlight the significance of customizing educational experiences to meet the needs of each individual student in the context of personalized learning. The model's excellent precision and accuracy suggest that it could be helpful in pinpointing particular areas in which pupils might need more help or enrichment. These findings can be used to inform the development of individualized learning strategies that address the individual strengths and limitations of each student, resulting in a more successful and inclusive learning environment.

In summary, even though the model shows good predictive accuracy and little variation in subject marks between sites, the substantial differential impact indicates that fairness and inclusion in the context of personalized learning need to be carefully considered. Ensuring the model's dependability and fair influence on students in Data Science education will require constant observation, continuous assessment, and proactive steps to overcome prejudices.

STRENGHTS AND WEAKNESSES OF THE STUDY

Strengths of the Study:

High Model Performance: The study shows that the model is successful in forecasting student performance, with a high degree of accuracy (97%) and precision across various classes. This powerful capacity to foretell the future is a core strength.

Comprehensive Evaluation Metrics: Detailed assessment measures, like the differential impact, categorization report, and confusion matrix, are used to provide a comprehensive knowledge of the model's performance from multiple perspectives. The thorough assessment raises the study's trustworthiness.

Feature Variability Analysis: The research gains depth by utilizing one-way ANOVA testing to evaluate the diversity of subject marks across several locations. This method assists in locating any possible regional differences in academic achievement.

Correlation Analysis: A more sophisticated understanding of the relationships between academic performance indicators is made possible by the analysis of correlations between subject marks and the display of correlation matrices. This offers insights into the independence of several subjects.

Attention to Fairness: The study's dedication to justice and inclusivity is demonstrated by the recognition and examination of the disparate impact. Acknowledging potential prejudices is an essential first step toward guaranteeing fair results for a variety of student demographics.

Consideration of Personalized Learning Implications: The study recognizes the potential for customizing educational experiences to meet the needs of individual students based on the insights provided by the model and analyzes how the findings relate to personalized learning. This gives the study a more useful perspective.

Weaknesses of the Study:

Limited Demographic Analysis: Although the study looks at location-based variability, academic performance may also be influenced by other demographic factors (such as age, financial situation, or status as an international student). A more thorough understanding of potential biases might be obtained through a more thorough demographic analysis.

Absence of Root Cause Analysis for Disparate Impact: The report notes a significant disparate impact but doesn't explore the underlying reasons behind this imbalance. If and why particular demographic groups are disproportionately affected, a more thorough investigation might reveal.

Potential Data Limitations: The dataset's representativeness and place of origin are not specifically covered in the study. Accurately interpreting the results requires an understanding of the source of the dataset and any potential biases included in it.

Lack of Mitigation Strategies: While the study identifies the disparate impact as a concern, it does not propose specific mitigation strategies or interventions to address potential biases. Suggesting practical steps for improvement would strengthen the study.

Timeline Chart:

TIMLINE CHART: Week 1: Planning & Preparation

Day 1-2: Outline the objectives and purpose of the study;

Day 3–4: Literature Review Gather and Review Relevant Works

Day 5–7: Develop Research Questions and Hypotheses

Week 2: Information Gathering Day 8–10: Develop data collection instruments (such as interview guides and questionnaires).

Day11–12: Obtain any necessary ethics permissions.

Day 13–14: Collect Data (via surveys, interviews, case studies, etc.)

Week 3: Reviewing the Data

Day 15–18: Conduct Quantitative Data Analysis (Predictive Modeling)

Day 19–21: Conduct Thematic and Content Analysis, among Other Qualitative Data Analysis Methods

Integrating Qualitative and Quantitative Findings on Day 22

Week 4: Presentation and Reporting

Day 23–25: Get the Research Findings and Discussion Ready

Day 26–27: Write a research report.

REFERENCES

1. R. Alamri and B. Alharbi, "Explainable Student Performance Prediction Models: A Systematic Review," in IEEE Access, vol. 9, pp. 33132-33143, 2021, doi: 10.1109/ACCESS.2021.3061368.

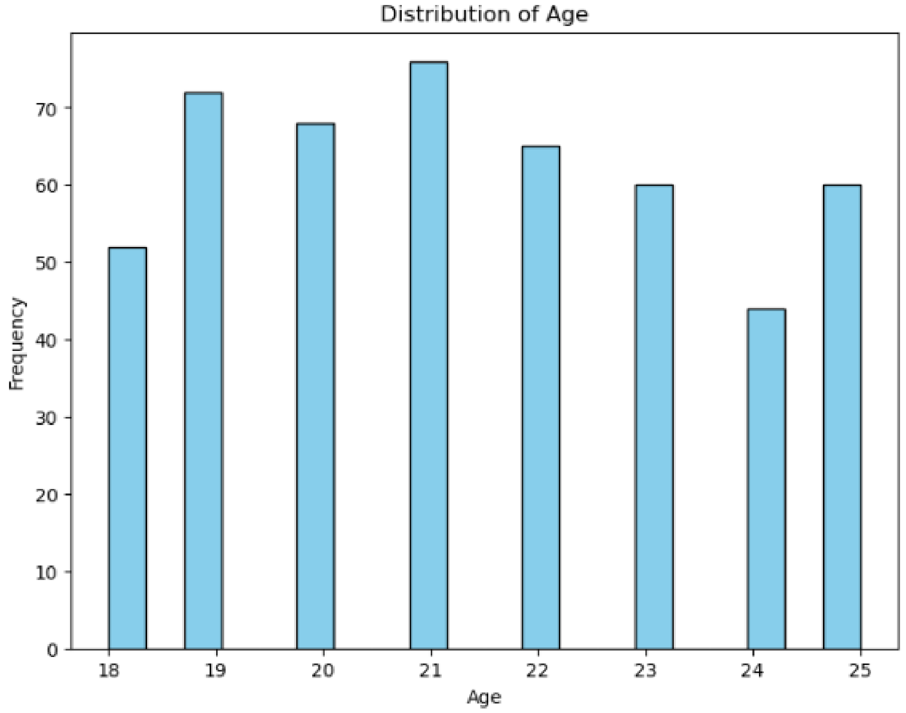
   keywords: {Predictive models;Systematics;Stakeholders;Numerical models;Context modeling;Taxonomy;Deep learning;Explainable artificial intelligence;explainable machine learning;student performance models;systematic literature review},

2. C. Huang, Z. Zhang, B. Mao and X. Yao, "An Overview of Artificial Intelligence Ethics," in IEEE Transactions on Artificial Intelligence, vol. 4, no. 4, pp. 799-819, Aug. 2023, doi: 10.1109/TAI.2022.3194503.

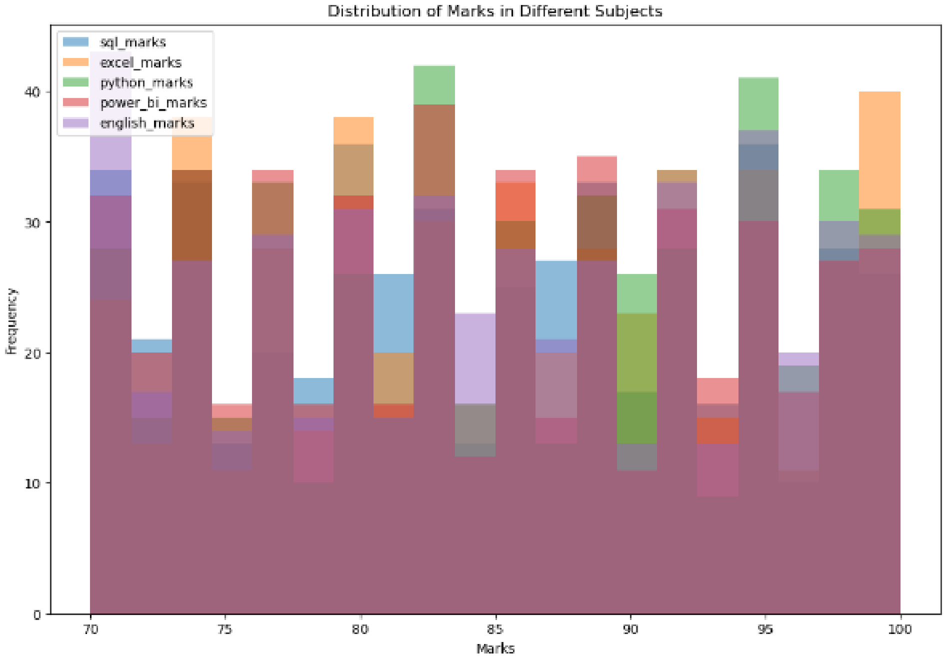
   keywords:{Artificialintelligence;Ethics;Guidelines;Privacy;Government;Systematics;Security;Artificial intelligence (AI);AI ethics;ethical issue;ethical theory;ethical principle},

APPENDIX:

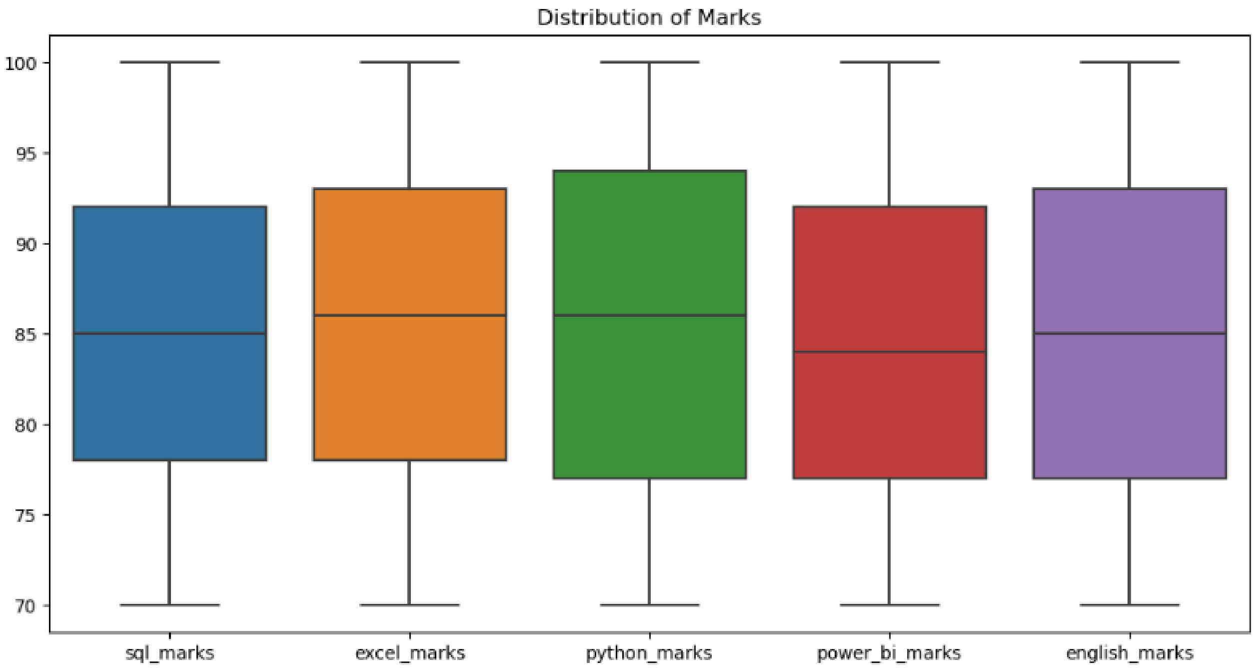
```
# Visualize the distribution of age
plt.figure(figsize=(8, 6))
plt.hist(df['age'], bins=20, color='skyblue', edgecolor='black')
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



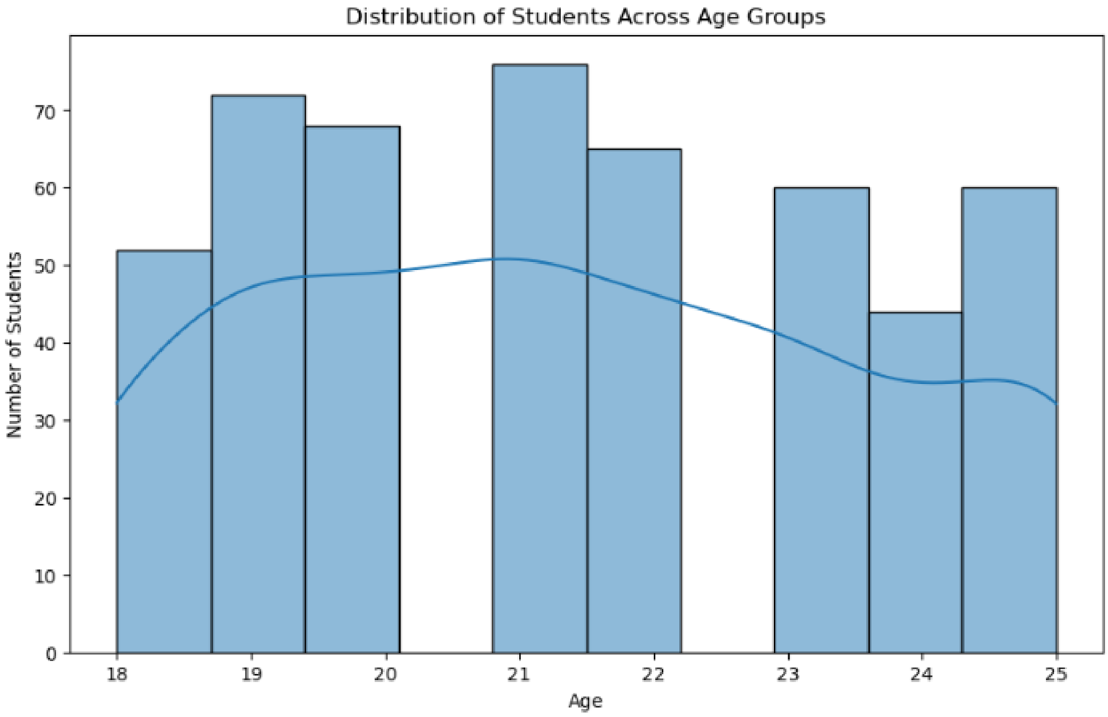
```
# Visualize the distribution of marks in different subjects
subjects = ['sql_marks', 'excel_marks', 'python_marks', 'power_bi_marks', 'english_marks']

plt.figure(figsize=(12, 8))
for subject in subjects:
    plt.hist(df[subject], bins=20, alpha=0.5, label=subject)

plt.title('Distribution of Marks in Different Subjects')
plt.xlabel('Marks')
plt.ylabel('Frequency')
plt.legend()
plt.show()
```




```
# Distribution of marks
plt.figure(figsize=(12, 6))
sns.boxplot(data=df.iloc[:, 3:])  # Considering columns from 'sql_marks' onwards
plt.title("Distribution of Marks")
plt.show()
```



```python
# Analyze the distribution of students across age groups
plt.figure(figsize=(10, 6))
sns.histplot(df['age'], bins=10, kde=True)
plt.title('Distribution of Students Across Age Groups')
plt.xlabel('Age')
plt.ylabel('Number of Students')
plt.show()
```




```python
print(f"Accuracy: {accuracy:.2f}")
print(f"Precision: {precision:.2f}")
print("\nConfusion Matrix:")
print(conf_matrix)
print("\nClassification Report:")
print(classification_rep)
```

```
Accuracy: 0.97
Precision: 0.97

Confusion Matrix:
[[ 7  0  1]
 [ 0 11  2]
 [ 0  0 79]]

Classification Report:
              precision    recall  f1-score   support

        High       1.00      0.88      0.93         8
         Low       1.00      0.85      0.92        13
      Medium       0.96      1.00      0.98        79

    accuracy                           0.97       100
   macro avg       0.99      0.91      0.94       100
weighted avg       0.97      0.97      0.97       100
```

```python
# Perform one-way ANOVA for each performance column across different locations
for col in performance_columns:
    print(f"\nOne-way ANOVA for {col} across different locations:")
    locations = df['location'].unique()
    groups = [df[df['location'] == loc][col] for loc in locations]
    f_statistic, p_value = f_oneway(*groups)
    print(f"F-statistic: {f_statistic}, p-value: {p_value}")
```

```
One-way ANOVA for sql_marks across different locations:
F-statistic: 1.6650385827848695, p-value: 0.10442875503791374

One-way ANOVA for excel_marks across different locations:
F-statistic: 0.4631681599530263, p-value: 0.8819627377483973

One-way ANOVA for python_marks across different locations:
F-statistic: 0.6080642226261233, p-value: 0.7713645050930659

One-way ANOVA for power_bi_marks across different locations:
F-statistic: 0.7170379070308298, p-value: 0.6765733669530619

One-way ANOVA for english_marks across different locations:
F-statistic: 0.8298612847521354, p-value: 0.5765238566489995
```

```python
disparate_impact = conf_matrix[1, 1] / (conf_matrix[0, 1] + 1e-10)
print(f"\nDisparate Impact: {disparate_impact}")
```

```
Disparate Impact: 110000000000.0
```

```python
# Plotting correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix of Academic Performance')
plt.show()
```