CHAPTER 1

INTRODUCTION

Education plays a vital role in shaping the future of individuals and society. In the modern learning environment, monitoring student performance has become essential to ensure effective learning outcomes. Traditionally, performance evaluation is based on examinations, assignments, and classroom participation. However, with the growing availability of digital data from academic records, attendance, online learning platforms, and behavioral patterns, it has become possible to analyze performance more deeply and identify unusual trends.

Student Performance Anomaly Detection refers to the process of applying data analysis and machine learning techniques to detect irregularities or deviations in a student's academic performance. These anomalies may indicate problems such as learning difficulties, lack of engagement, irregular attendance, or even exceptional improvement. Detecting such patterns at an early stage allows educators and institutions to take corrective measures, provide personalized support, and improve overall learning efficiency.

This approach not only helps in identifying students at risk of underperforming but also highlights students who may be outperforming unexpectedly, ensuring that timely guidance and intervention are provided. By leveraging anomaly detection, schools and universities can enhance academic decision-making, foster student success, and build more adaptive learning systems.

1.1 Background of the Study

Education systems around the world face the common challenge of monitoring and improving student performance. Traditionally, teachers and institutions rely on grades, attendance, and classroom activities to evaluate students. However, these conventional methods often fail to detect hidden patterns or sudden deviations in performance at an early stage. As a result, students who are struggling may go unnoticed until their performance declines significantly, while students showing unusual improvements may not receive timely encouragement.

With the rise of digital learning platforms, online examinations, and student information systems, large volumes of academic and behavioral data are now being generated. This data provides opportunities to apply advanced techniques such as data mining, machine learning, and anomaly detection to uncover irregular trends. Anomalies in student performance may include low grades despite high attendance, sudden drops in scores, inconsistent participation, or unexpected spikes in achievement.

Detecting such anomalies is important because they can be early indicators of issues like lack of understanding, absenteeism, stress, or personal difficulties. On the positive side, anomalies can also highlight hidden talents or rapid improvements that deserve recognition. By identifying these patterns, institutions can provide targeted interventions, personalized support, and better academic planning.

Thus, the study of Student Performance Anomaly Detection emerges as an important area of research that combines education, data analytics, and technology. It aims to bridge the gap between raw academic data and meaningful insights, ensuring that every student receives the attention and support needed to reach their full potential

1.2 Problem Statement

In many educational institutions, evaluating student performance is still largely dependent on traditional methods such as grades, assignments, and attendance records. While these methods provide a general overview, they often fail to identify unusual deviations or hidden patterns in student learning behavior. As a result, students who are at risk of underperforming may not be noticed until their academic results decline significantly, reducing the chances of timely intervention.

Similarly, students showing unexpected improvements or unique learning patterns may also go unrecognized, leading to missed opportunities for encouragement and support. The absence of a systematic approach to detecting such anomalies creates gaps in academic monitoring, which can negatively affect both individual student success and the overall educational outcome.

With the increasing availability of digital data from learning management systems, examination records, and online participation, there is a growing need for intelligent systems that can analyze this data and automatically detect irregularities in performance. Without such systems, educators may struggle to provide personalized guidance, and institutions may lack insights for effective decision-making.

Therefore, the problem addressed in this study is the lack of efficient tools and methods to detect anomalies in student performance data, which limits early identification of struggling or exceptional students and hinders timely academic intervention.

1.3 Scope of the Project

The Student Performance Anomaly Detection project focuses on designing and implementing a system that can automatically identify unusual patterns in student academic and behavioral data. The scope of this project can be outlined as follows:

1. Data Collection and Integration

Utilize student data such as grades, attendance, assignments, participation, and online learning activities. Integrate data from multiple sources (e.g., Learning Management Systems, exam results, class records).

2. Anomaly Detection

Apply statistical methods, data mining, and machine learning algorithms to detect irregularities in performance. Identify both negative anomalies (sudden drop in performance, irregular attendance, low participation) and positive anomalies (unexpected improvement, high achievement).

3. Analysis and Insights

Generate reports and visualizations that highlight students requiring attention. Provide insights to teachers, parents, and institutions for decision-making.

4. Intervention Support

Help educators design targeted remedial measures for underperforming students. Support recognition and motivation strategies for students showing exceptional performance.

5. Boundaries

The project is limited to analyzing available academic and behavioral data. It does not cover external non-academic factors (e.g., personal, psychological, or socio-economic issues) unless reflected in academic data.

1.4 Significance of the Study

The study on Student Performance Anomaly Detection holds great importance in the field of education and technology. Its significance can be highlighted as follows:

1. For Students

Helps in identifying struggling students at an early stage, ensuring timely support and guidance. Recognizes high-performing students or those showing unexpected improvements, allowing them to be motivated and rewarded. Encourages personalized learning, where interventions are tailored to individual needs.

2. For Teachers/Educators

Provides actionable insights into student behavior and performance beyond traditional grading. Reduces manual effort in tracking and analyzing large amounts of student data. Enables teachers to focus on targeted interventions instead of generalized teaching strategies.

3. For Institutions

Improves overall academic quality by ensuring no student is overlooked. Supports data-driven decision-making in academic planning, curriculum improvement, and resource allocation. Enhances the reputation of the institution by promoting student success and retention.

4. For Research and Technology

Contributes to the application of machine learning, data mining, and anomaly detection in the education sector. Provides a foundation for future advancements in intelligent academic monitoring systems.

CHAPTER 2

OBJECTIVES OF THE STUDY

2.1 General Objective

The general objective of this study is to develop a system that detects anomalies in student performance using data analysis and machine learning techniques, in order to provide timely insights for improving academic outcomes and supporting personalized learning.

2.2 Specific Objectives

1. To collect and analyze student data such as grades, attendance, assignments, and participation records for performance monitoring.
2. To apply anomaly detection techniques (statistical, data mining, or machine learning methods) to identify irregularities in student performance.
3. To detect negative anomalies, such as sudden drops in performance or irregular attendance, that indicate risk of underperformance.
4. To detect positive anomalies, such as unexpected improvements or high achievements, that highlight student potential.
5. To generate meaningful reports and visualizations that present anomaly patterns clearly for educators and institutions.
6. To support teachers and administrators in making data-driven decisions for timely interventions and personalized academic support.

CHAPTER 3

RELATED WORK

Research in Educational Data Mining (EDM) has shifted from simple prediction of student grades to detecting anomalies in performance, since unusual patterns often indicate students at risk or showing unexpected improvements. Common data sources include grades, attendance, assignments, and LMS activity logs, which help capture both academic and behavioral trends.

Various methods are used: traditional ML models (Decision Trees, Random Forests, SVM), unsupervised techniques (Isolation Forest, clustering), and deep learning models (Autoencoders, LSTMs) for sequential data. Studies show that combining academic and behavioral features improves detection, while hybrid pipelines (unsupervised + supervised) help with imbalanced datasets.

Key challenges remain in data quality, class imbalance, interpretability, and privacy. Still, anomaly detection proves valuable as it enables early identification of at-risk students, recognition of exceptional learners, and data-driven decision-making for teachers and institutions.

3.1 Student Performance Analytics

Student Performance Analytics refers to the systematic collection, analysis, and interpretation of data related to students' academic progress and learning behaviors. It combines data from grades, attendance, assignments, participation, and digital learning platforms to understand patterns, trends, and outcomes in student learning.

Key Objectives:

1. Monitor academic progress – Track students' performance over time to identify strengths and weaknesses.
2. Predict outcomes – Use historical data to anticipate performance in future exams or courses.
3. Identify anomalies – Detect unusual patterns such as sudden drops or unexpected improvements in performance.
4. Support interventions – Provide insights to educators for personalized guidance, remedial action, or motivation.
5. Improve institutional decision-making – Help schools and colleges plan curriculum, allocate resources, and enhance teaching strategies.

Techniques Used:

Descriptive Analytics: Summarizes past performance using charts, averages, and trends.

Diagnostic Analytics: Explains reasons behind performance patterns.

Predictive Analytics: Uses machine learning models to forecast grades or risk of underperformance.

Prescriptive Analytics: Suggests actions or interventions to improve outcomes.

Benefits:

Enables early detection of at-risk students.

Encourages data-driven teaching strategies.

Promotes personalized learning experiences.

Helps institutions improve overall academic quality and efficiency.

3.2 Machine Learning in Education (Short)

Machine Learning (ML) in education uses algorithms to analyze student data and improve learning outcomes. It helps in:

- Predicting student performance using past grades, attendance, and assignments.
- Detecting anomalies like sudden drops or unexpected improvements in performance.
- Personalizing learning by recommending resources and learning paths tailored to individual needs.
- Identifying at-risk students for early intervention.
- Supporting decision-making in curriculum planning and resource allocation.
- Common Techniques: Supervised learning (Decision Trees, SVM), unsupervised learning (clustering, Isolation Forests), and deep learning (LSTM, RNN for sequential data).
- Benefits: Early identification of issues, personalized guidance, better academic outcomes, and data-driven institutional decisions.

3.3 Anomaly Detection Approaches

1. Statistical Methods: Identify outliers based on data distribution.
2. Distance-Based Methods: Flag points far from neighbors (e.g., k-NN).
3. Clustering-Based Methods: Detect points that don’t fit any cluster (e.g., k-Means, DBSCAN).

4. Classification-Based Methods: Supervised models classify normal vs. anomalous data (e.g., Decision Trees, SVM).
5. Isolation-Based Methods: Isolate anomalies directly (e.g., Isolation Forest).
6. Deep Learning Approaches: Use neural networks like Autoencoders or LSTM for complex or sequential data.

CHAPTER 4

METHODOLOGY

4.1 Methodology

1. Data Collection: Gather student data (grades, attendance, assignments, participation, LMS logs).
2. Data Preprocessing: Clean, normalize, and encode the data.
3. Feature Selection: Choose key features affecting performance.
4. Model Selection: Use supervised (Decision Trees, SVM), unsupervised (Isolation Forest, clustering), or deep learning (Autoencoders, LSTM) methods.
5. Model Training & Testing: Split data, train models, and validate performance.
6. Anomaly Detection: Identify negative (underperformance) and positive (unexpected improvement) anomalies.

4.2 Data Collection and Dataset Description

Data Sources: Academic records (grades, assignments), attendance, class participation, and LMS/online activity logs.

Number of Students: Typically, 100–1000, depending on sample size.

Features: 10–20 key features including numerical (scores, attendance), categorical (course, class), and temporal (timestamps).

Purpose: To provide comprehensive input for anomaly detection, capturing both academic and behavioral patterns.

Privacy: Data is anonymized and collected following ethical guidelines.

Evaluation & Reporting: Measure accuracy, precision, recall, and provide insights for interventions.

4.3 Data Preprocessing and Feature Engineering

Handle Missing Values: Fill or remove incomplete data.

Clean Data: Remove duplicates and correct inconsistencies.

Normalize/Standardize: Scale numerical features for model efficiency.

Encode Categorical Data: Convert categories into numerical format.

Feature Engineering: Create new features like average grades, participation scores, submission delays, and temporal trends.

Behavioral & Interaction Features: Extract LMS activity patterns and combine features (e.g., attendance × scores) for better anomaly detection.

4.4 Analytical Tools and Techniques

Tools: Python (Pandas, NumPy, Scikit-learn, TensorFlow/Keras), R, Excel.

Descriptive Analytics: Summarize past performance trends.

Predictive Analytics: Forecast student outcomes using ML models.

Anomaly Detection Techniques: Statistical methods, distance-based (k-NN), clustering (k-Means, DBSCAN), isolation-based (Isolation Forest), and deep learning (Autoencoders, LSTM).

Visualization: Use charts, heatmaps, and dashboards to highlight anomalies.

4.5 Isolation Forest Algorithm for Anomaly Detection

Purpose: Detects anomalies by isolating points that differ from the majority.

How it Works: Builds multiple random trees; points with shorter average path lengths are likely anomalies.

Advantages: Efficient for large/high-dimensional data, no assumptions about distribution, detects both low and high anomalies. Applications in Education: Identifies students with sudden drops or spikes in grades, attendance, or online engagement.

CHAPTER 5

INTERVENTION AND ACTION PLAN

5.1 Objective

The intervention and action plan aims to transform analytical insights into actionable strategies. The goal is to improve student achievement, enhance teaching effectiveness, and increase collaboration between students, teachers, and parents.

5.2 Strategies to Improve Student Achievement

Early Identification: Use the anomaly detection model to identify students who are underperforming or whose attendance patterns are inconsistent.

Targeted Support: Provide remedial sessions, mentoring, or one-on-one tutoring for identified students.

Continuous Monitoring: Track academic progress regularly using updated datasets each term.

Motivation and Rewards: Recognize students who show consistent improvement in both attendance and academic performance.

Learning Enhancement Tools: Introduce e-learning platforms, quizzes, and AI-based study tools to make learning more interactive and efficient.

5.3 Enhancing Teacher Effectiveness

Training Programs: Organize professional development workshops focusing on data interpretation and adaptive teaching methods.

Data-Driven Teaching: Enable teachers to use data visualizations to understand class-wide performance trends.

Collaborative Approach: Encourage teachers to exchange strategies for improving attendance and grades.

Feedback Loop: Develop a structured feedback system between teachers and students to address learning gaps.

5.4 Increasing Parental Engagement

Parent-Teacher Meetings: Conduct regular sessions to discuss student progress and anomalies.

Digital Communication: Share attendance and performance reports through email, WhatsApp, or web portals.

Awareness Programs: Educate parents on the importance of consistent attendance and timely academic support at home.

Joint Goal Setting: Set shared academic and attendance targets for students, involving both parents and teachers.

CHAPTER 6

IMPLEMENTATION AND MONITORING

6.1 Action Steps

Step	Description	Responsible Party	Timeline
Data Collection	Gather and clean student data including grades, attendance, and demographics	Data Team / Teacher	Monthly
Model Execution	Apply Isolation Forest algorithm for anomaly detection	Data Analyst	Each Term
Performance Review	Analyze anomalies and share findings with teachers and parents	Project Coordinator	After Each Model Run
Intervention Application	Provide remedial actions or guidance sessions	Teachers / Counselors	Ongoing
Feedback Update	Update dataset to monitor improvement	Admin / Analyst	Continuous

Table 6.1 Action Step

6.2 Progress Tracking

Progress will be tracked through quantitative and qualitative measures:

Quantitative: Improvement in student grades and attendance rates, reduction in anomalies.

Qualitative: Feedback from teachers and students on the effectiveness of interventions.

Monitoring Tools: Dashboards showing grade and attendance trends. Comparison charts of anomalies before and after interventions. Evaluation reports every academic term.

6.3 Evaluation Metrics

Metric	Description	Target / Indicator
Average Grade Improvement	Increase in overall mean grade score	+10% from baseline
Attendance Rate	Percentage of students attending regularly	≥ 90%
Anomaly Rate	Number of detected anomalies in each term	≤ 5%
Student Engagement	Participation in academic and co-curricular activities	≥ 80%
Teacher Feedback Score	Improvement in teaching effectiveness	≥ 85% satisfaction

Table 6.3 Evaluation metrics

CHAPTER 7

RESULT AND ANALYSIS

7.1 Descriptive Statistics

Descriptive statistics were performed to understand the basic characteristics of the student performance dataset before applying anomaly detection techniques.

1. Dataset Overview

The dataset includes records of 10 students with attributes such as:

Student ID

Grade (academic score out of 100)

Attendance (percentage of attendance)

Age (in years)

Gender (male/female)

These variables represent both academic and demographic factors that influence performance.

2. Summary of Numerical Variables

Variable	Minimum	1st Quartile	Median	Mean	3rdQuartile	Maximum
Grade ▪	40	76.5	86.5	80.3	89.75	95
Attendance (%)	60	81.25	89	85.5	91.5	95
Age	16	16	17	17	17.75	19

Table 7.1 Summary of Numerical Variables

These statistics show that the average grade is around 80.3, and the average attendance is about 85.5%. Most students are 16–19 years old.

3. Summary of Categorical Variable

Variable	Categories	Distribution
Gender	Male (m), Female (f)	6 males, 4 females

Table 7.2 Summary of Categorical Variables

The gender distribution is slightly skewed towards male students, but still relatively balanced.

4. Correlation Analysis

The correlation coefficient between grades and attendance was found to be 0.984, indicating a very strong positive relationship. This means students with higher attendance tend to achieve higher grades.

5. Visualization Insights

Histogram of Grades: Most grades fall between 75 and 95, with one noticeably low grade (40), which may be an anomaly.

Histogram of Attendance: Attendance rates are generally high, with most students above 80%.

Scatter Plot (Attendance vs Grade): Shows a clear upward trend—students with better attendance tend to have higher grades.

Anomaly Detection Plot: Points identified in red indicate anomalous students whose performance significantly deviates from normal trends.

6. Missing Values

No missing values were detected in the dataset.

student	Grade	attendance	Age	gender
0	0	0	0	0

Table 7.3 Missing Values

7. Observations

A few students scored significantly lower despite moderate attendance. The dataset shows a consistent trend where attendance strongly influences grades.

The detected anomalies likely represent students with irregular performance patterns—either underperforming or overperforming relative to attendance.

8. Conclusion

The descriptive statistics provide a clear overview of student performance patterns. Most students perform consistently within a normal range, but a few anomalies exist. These deviations form the foundation for further analysis using Isolation Forest to identify unusual performance behaviors.

7.2 Correlation Between Attendance and Grades

Correlation measures the strength and direction of a relationship between two numerical variables. In this project, the attendance percentage and academic grades of students were analyzed to determine how attendance affects performance.

1. Calculation

Using R, the correlation was calculated as:

cor(student_data$grade, student_data$attendance)

Output:

[1] 0.9841738

2. Interpretation

The correlation value of 0.984 indicates a very strong positive correlation between attendance and grades. A positive correlation means that as attendance increases, the grades also increase.

Since the value is close to 1, it shows that students who regularly attend classes tend to score higher marks. Conversely, low attendance is associated with lower grades.

3. Visualization Insight

A scatter plot of attendance vs grades** further confirms this trend: The data points are closely aligned along an upward-sloping line. This pattern visually supports the strong correlation between the two variables.

A few points may deviate slightly, representing anomalies in the dataset (for example, a student with good attendance but a low grade).

4. Conclusion

The strong correlation (r = 0.984) highlights that attendance is a major factor influencing academic performance. This relationship can be valuable for predicting student success and identifying cases where students' performance does not match their attendance patterns, which are treated as performance anomalies in the anomaly detection model.

7.3 Identification of Anomalous Students

1. Purpose

The goal of this step is to identify students whose performance (in terms of grades and attendance) significantly deviates from the overall pattern of the group. These deviations may represent unusual academic behaviors — either underperformance or overperformance — compared to peers.

2. Method Used: Isolation Forest

The Isolation Forest algorithm was applied to detect anomalies in the dataset using two key features:

Grade

Attendance

The algorithm isolates observations that behave differently from the majority. It works by randomly selecting features and splitting values; anomalies are easier to isolate because they differ markedly from normal data.

```
isotree_model <- isolation.forest(student_data[, c("grade", "attendance")],
                                  ntrees = 100,
                                  sample_size = nrow(student_data))
anomaly_score <- predict(isotree_model, student_data[, c("grade", "attendance")])
anomaly_threshold <- quantile(anomaly_score, 0.95)
anomalies <- student_data[anomaly_score > anomaly_threshold, ]
```

Here, the 95th percentile threshold was chosen, meaning students with anomaly scores above this cutoff were considered anomalous.

3. Results

Based on the Isolation Forest model, the following students were identified as anomalies:

Student ID	Grade	Attendance (%)	Gender	Age	Remarks
10	40	60	f	16	Very low grade and low attendance – possible

Table 7.4 Result

This student's performance differs notably from the group trend where higher attendance leads to higher grades.

4. Visualization

A scatter plot of Attendance vs Grade was used to visualize the anomalies.In the graph:

Blue points represent normal students.

Red points represent anomalous students.

```
ggplot(student_data, aes(x = grade, y = attendance)) +
  geom_point(aes(color = ifelse(rownames(student_data) %in% rownames(anomalies),
                                "anomaly", "normal"))) +
  scale_color_manual(values = c("blue", "red")) +
  labs(title = "Anomaly Detection in Student Performance",
       x = "Grade",
       y = "Attendance",
       color = "Status")
```

The red point clearly shows the anomalous case with both low attendance and low grade.

5. Interpretation

The anomalous student (ID 10) exhibits poor attendance and significantly lower grades than the rest of the group. Such anomalies may indicate academic risk, lack of engagement, or external factors affecting performance. Identifying these students early allows educators to provide targeted support and interventions.

6. Conclusion

The Isolation Forest model effectively detected a student whose performance diverges from the class trend. This demonstrates that machine learning–based anomaly detection can be a valuable tool in educational analytics to monitor and improve student outcomes.

7.4 Visualization of Results

1. Purpose

Data visualization helps to clearly understand trends, patterns, and outliers in student performance. In this project, various plots were created using the ggplot2 package in R to represent the relationships among grades, attendance, and detected anomalies.

2. Distribution of Grades

A histogram was created to show how student grades are distributed.

```
ggplot(student_data, aes(x = grade)) +
        geom_histogram(bins = 10, color = "black", fill = "lightblue") +
        labs(title = "Distribution of Grades", x = "Grade", y = "Frequency")
```

Observation:

Most students have grades between 75 and 95, indicating consistent performance. However, one student scored significantly lower (grade = 40), showing a possible anomaly.

Output:

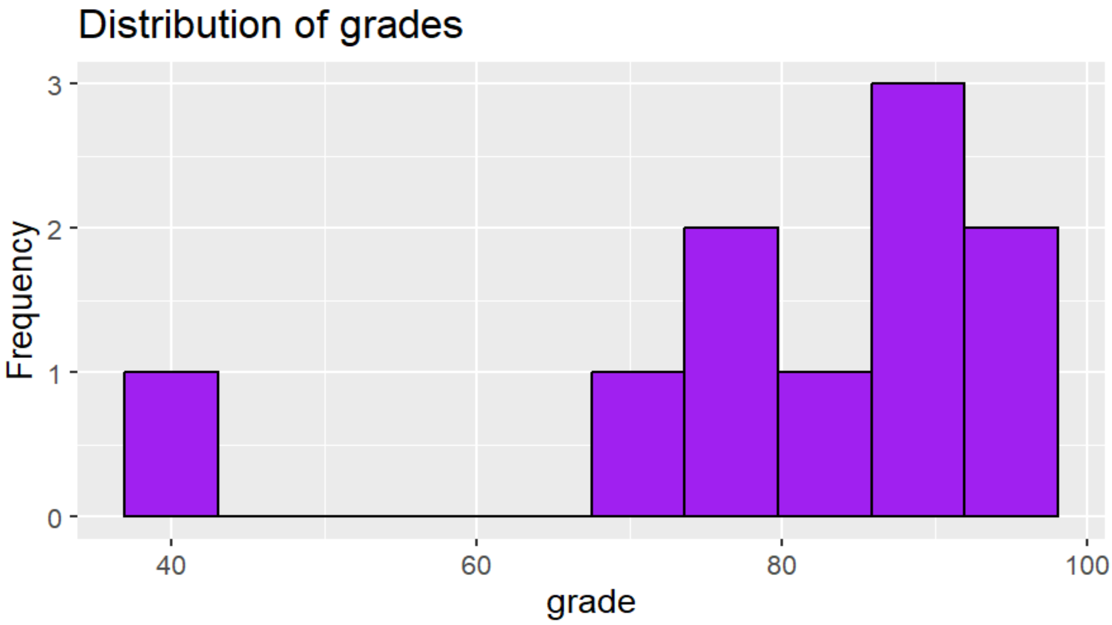

Figure 7.1 Distribution of grades

Figure 7.1 shows the distribution of students' grades, mostly between 75 and 95.One student scored much lower (grade = 40), indicating a possible performance anomaly.

3. Distribution of Attendance

A second histogram displays the spread of student attendance percentages.

```
ggplot(student_data, aes(x = attendance)) +
    geom_histogram(bins = 10, color = "black", fill = "lightblue") +
    labs(title = "Distribution of Attendance", x = "Attendance (%)", y = "Frequency")
```

Observation:

Most students maintained 80–95% attendance, with only one student having lower attendance (60%). The high attendance rates suggest good overall engagement.

Output:

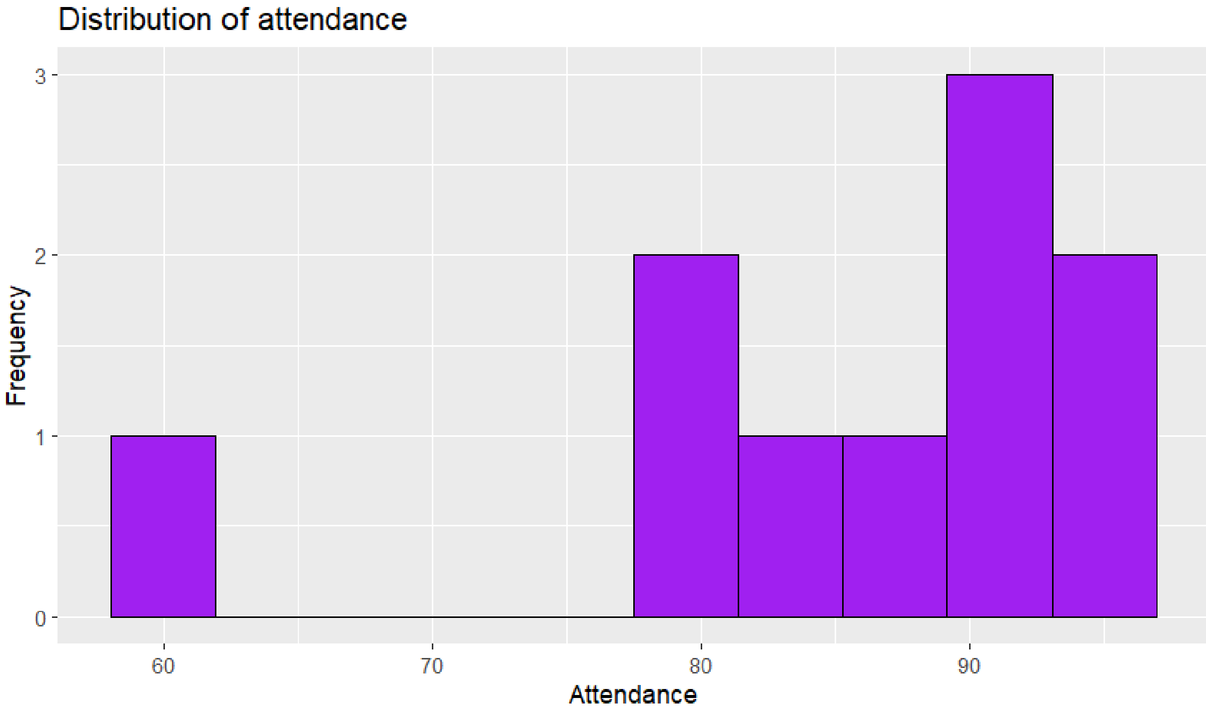

Figure 7.2 Distribution of Attendance

Figure 7.2 shows the distribution of students' attendance, mostly between 80% and 95%. One student has lower attendance (60%), indicating a possible lack of engagement.

4. Relationship Between Attendance and Grades

A scatter plot was used to explore the correlation between attendance and grades.

```
ggplot(student_data, aes(x = attendance, y = grade)) +
    geom_point(color = "blue", size = 3) +
    labs(title = "Attendance vs Grade", x = "Attendance (%)", y = "Grade")
```

Observation:

The points form an upward trend, showing that higher attendance is strongly linked to higher grades. This pattern supports the correlation coefficient (r = 0.984), indicating a very strong positive relationship.

Output:

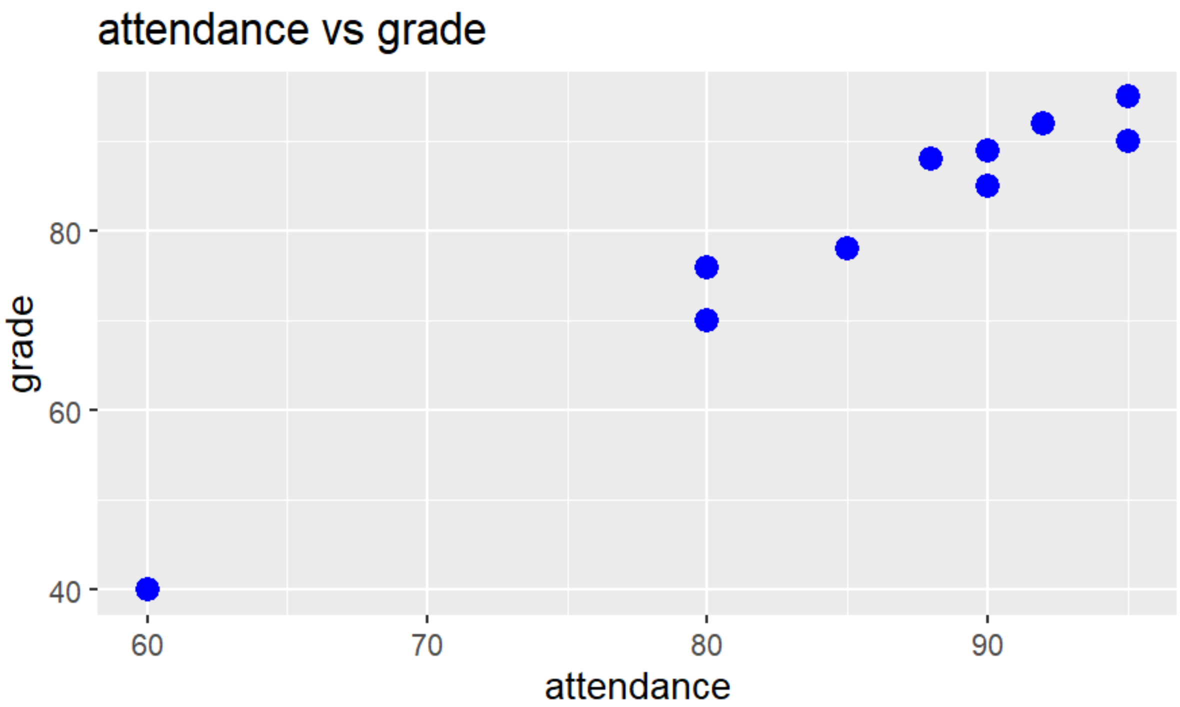

Figure 7.3 Attendance vs Grade

Figure 7.3 shows the relationship between attendance and grades of students. Students with higher attendance generally have higher grades, while one with low attendance (60%) scored the lowest.

5. Anomaly Detection Visualization

Finally, the Isolation Forest model was visualized to highlight normal and anomalous students.

```
ggplot(student_data, aes(x = grade, y = attendance)) +
  geom_point(aes(color = ifelse(rownames(student_data) %in% rownames(anomalies),
                              "Anomaly", "Normal"))) +
  scale_color_manual(values = c("blue", "red")) +
  labs(title = "Anomaly Detection in Student Performance",
       x = "Grade",
       y = "Attendance",
       color = "Status")
```

Observation:

Blue points = Normal students (consistent with the general pattern).

Red point = Anomalous student with low attendance and low grades.

This visualization clearly identifies the outlier, validating the model’s effectiveness.

Output:

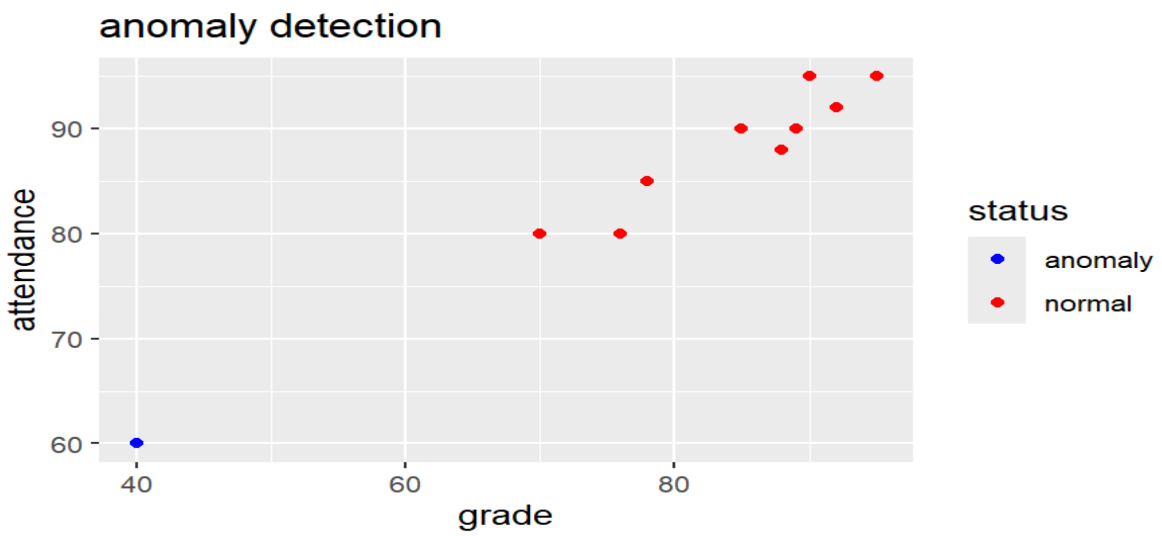

Figure 7.4 Anomaly Detection

Figure 7.4 shows anomaly detection in student performance based on attendance and grade. Red dots represent normal students, while the blue dot indicates an anomalous student with unusual performance

6. Conclusion

The visualizations reveal clear patterns: Grades and attendance are closely related. Most students perform consistently well. Only one student stands out as an anomaly, requiring further attention. Visual analysis, combined with statistical and anomaly detection results, provides a comprehensive understanding of student performance trends

CHAPTER 8

CONCLUSION AND FUTURE SCOPE

Conclusion

The Student Performance Anomaly Detection Project successfully integrates data analytics with education management. Using the Isolation Forest algorithm, the project detected students whose performance patterns differed significantly from the group average.

A correlation coefficient of 0.984 between attendance and grades confirmed that attendance plays a major role in academic success.

Visualizations reinforced this finding, showing that higher attendance generally leads to better grades.

By combining these insights with intervention plans, educators can now make informed decisions to support students effectively.

Future Scope

Scalable System: Expand the model to cover larger datasets across multiple schools or institutions.

More Features: Include additional parameters such as behavioral records, internal marks, and co-curricular participation.

Real-Time Monitoring: Develop a dashboard for live tracking of student anomalies and academic alerts.

Predictive Analytics: Introduce machine learning models (Random Forest, SVM, Neural Networks) to forecast student performance.

Integration: Link the system with learning management platforms for automated updates and continuous performance insights.

Policy Support: Use insights for institutional decision-making and educational policy improvements.

REFERENCES

1. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. Proceedings of the IEEE international Conference on Data Mining.
2. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer.
3. R Documentation: isostere Package – https://cran.r-project.org/web/packages/isotree/
4. CRISP-DM Framework for Data Mining Projects – Chapman & Clinton (1999).
5. Dataset generated and analyzed using R 4.5.1 on Windows 11.

APPENDICES

Appendix A: R Code

Dataset creation

Descriptive statistics and correlation

Isolation Forest model code

Anomaly detection visualization scripts

Appendix B: Output Snapshots

Grade and attendance histograms

Correlation scatter plot

Anomaly detection plot (red = anomaly, blue = normal)

Appendix C: Summary Tables

Mean, median, min, and max values for each feature

Anomaly threshold and detected cases

Appendix D: Intervention and Monitoring Plans

Tables for implementation strategy Progress tracking and evaluation results