Dr Preetha V, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Determining the Statistical measures on student's Performance Metrics using Python code

Dr Preetha V
Assistant Professor of Computer Science, Sri S.Ramasamy Naidu College, Sattur

Abstract- This research investigates the student's performance metrics using statistical measures based on the dataset. Students studies and achievements have been distracted by various factors such as sleeping hours, internet access, motivational level and other environmental factors. The study of this research will be helpful to find the exact reasons for the performance of the students. Further, this research focusses on the python coding using pandas library to find the statistical measures. The research explored various datasets and find the best suitable dataset with 6000 entries for the exact analysis.

Keywords: Statistical measures, Student's Performance, Python,Datasets,Pandas.

I. INTRODUCTION

Academic performance is a multidimensional perspective influenced by school, College, University home, and personal habits. Educators often struggle to identify "at-risk" students early enough to intervene effectively and to improve their performance.The objective of the research was to quantify the impact of various independent variables on the dependent variable. Also, the statistical measures using Python code will help to analyse the large datasets satisfying the requirement of prediction, classification and other analysis.

Data Analysis:
Data analysis is a systematic process used to examine, clean, transform, and model data with the objective of discovering valuable information, drawing conclusions, and supporting decision-making. In the modern digital era, vast amounts of data are generated daily, making data analysis an essential discipline across fields such as business, science, healthcare, and technology.

Types of Data Analysis
Descriptive Analysis: It Focuses on summarizing historical data to understand what has happened.
Diagnostic Analysis
This type of analysis examines data to determine causes of past outcomes.
Predictive Analysis
This type of analysis Uses statistical models and machine learning to forecast future events.

Prescriptive Analysis
This type of analysis recommends actions based on analytical insights.

Core Concept of Data Analysis
Data by itself does not provide much value unless it is properly processed and understood. When raw data is carefully analyzed, it can be turned into meaningful insights. By using different analytical methods, it becomes possible to identify patterns, relationships, and trends within the data. These insights help both organizations and individuals make better and more informed decisions. Data analysis is not limited to a single field; it combines ideas from statistics, mathematics, and computer science. To carry out analysis effectively, tools such as Python, R, and SQL are widely used because they make handling and processing large amounts of data easier and faster.

Process of Data Analysis
The process of data analysis generally involves several important steps:

1. Data Collection
The first step is gathering data from various sources like databases, surveys, or online platforms. The quality of analysis depends heavily on the quality of collected data.

2. Data Cleaning

Once the data is collected, it often contains errors, missing values, or inconsistencies. These issues are corrected in this step to make the data reliable.

3. Data Transformation

After cleaning, the data is organized and converted into a format that is suitable for analysis. This may include sorting, filtering, or restructuring the data.

4. Data Modeling and Analysis

In this stage, different statistical or computational techniques are applied to study the data and extract useful information.

5. Data Visualization

The analyzed data is then presented in the form of charts, graphs, or dashboards. This makes it easier to understand complex information.

6. Interpretation

Finally, the results are interpreted to draw conclusions. These conclusions help in making decisions or solving problems effectively.

Statistical measures using Python

Python is a high-level, interpreted language that emphasizes code simplicity, making it the top choice for beginners and professionals alike across fields like web development, data science, and automation.For statistical analysis, pandas and numpy are essential for data manipulation. Descriptive statistics and Inference statistics can be easily perfomed in Pandas Python code.The built in support functions will ease the python code to obtain better results. To move observation to conclusion, relationships between variables and to find the correlation among variables, Python code is helpful for statisticians and for machine learning researches.

Existing Literature Review in Student's Peformance:

James S. Coleman and colleagues (1966) highlighted that family environment and access to resources play a significant role in shaping students' performance. Later research expanded to psychological and behavioral factors. Barry J. Zimmerman (2002) emphasized the importance of self-regulated learning, where students actively manage their own learning processes. With the advancement of technology, data-driven approaches have become increasingly important. Cristóbal Romero and Sebastián Ventura (2010) introduced educational data mining as a method to analyze student data and predict academic outcomes. In addition, learning analytics has gained attention as a tool for improving education systems. George Siemens (2013) discussed how analyzing large-scale educational data can help institutions design better learning environments and provide personalized support to students. Overall, the literature indicates that students' performance is influenced by a combination of personal, social, and institutional factors. Modern data analysis techniques provide powerful ways to understand these influences and improve educational outcomes.

Proposed work

The research was initiated with the data collection from kaggle recent datasets with the factors affecting student's performance.

Hours_Studied	Attendance	Extracurricular_Activities	Sleep_Hours	Previous_Scores	Motivation_Level	Internet_Access	Peer_Influence	Physical_Activity	Learning_Disabilities	Gender	Exam_Scores
23	84	No	7	73	Low	Yes	Positive	3	No	Male	67
19	64	No	8	59	Low	Yes	Negative	4	No	Female	61
24	98	Yes	7	91	Medium	Yes	Neutral	4	No	Male	74
29	89	Yes	8	98	Medium	Yes	Negative	4	No	Male	71
19	92	Yes	6	65	Medium	Yes	Neutral	4	No	Female	70

Figure 1. Student's Dataset

The performance metrics depends on the exam scores values.The factors chosen are Hours_studied, Attendance, Extra curricular activities of students,Internet access, Peer group influence,learning disabilities,motivational level,physical activity.

```
print("Total Records:", total_records)
print("Mean of Previous Scores:", mean_previous_scores)
print("Mean of Exam Scores:", mean_exam_scores)
print("Standard Deviation of Exam Scores:", sd_exam_scores)

print("\nMode and Count:")
print("Internet Access:", mode_internet, "| Count:", count_internet)
print("Motivation Level:", mode_motivation, "| Count:", count_motivation)
print("Learning Disabilities:", mode_learning, "| Count:", count_learning)
print("Peer Influence:", mode_peer, "| Count:", count_peer)
```

Figure 2. Python code

For mean calculations, we dropped the null values.

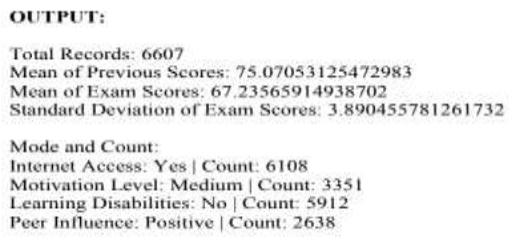

Figure 3. Output of statistical measures

The dataset contains 6607 records with both male and female data. It has been surveyed based on the influenced factors affecting exam scores. When compared with the previous exam scores, the current exam scores mean showed a downfall. When we analyse the factors, it was observed 6108 records were answered 'yes' for the Internet access factor.Hence, we conclude the above statistical measures mean, mode, standard deviation helps to determine the inference statistics.

II. CONCLUSION

This research paper analysis the performance of K-NN and Naïve bayes in the Iris dataset and implemented the results.It shows that both classifiers perfomed well with the training and testing datasets.
]
Conclusion
The research also highlights the most threatening factor "Internet Access" among students which marks a downfall in the exam scores. This will also affect the student's other performance achievements.

REFERENCES

1. Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). Equality of educational opportunity. U.S. Government Printing Office.
2. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 40(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532
3. Siemens, G. (2013). Learning analytics: The emergence of a discipline. American Behavioral Scientist, 57(10), 1380–1400. https://doi.org/10.1177/0002764213498851
4. Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2
5. Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). SAGE Publications.
6. Downey, A. (2014). Think stats: Exploratory data analysis in Python (2nd ed.). O'Reilly Media.
7. Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). SAGE Publications.
8. Hanushek, E. A., & Woessmann, L. (2010). The economics of international differences in educational achievement. Handbook of the Economics of Education, 3, 89–200. https://doi.org/10.1016/B978-0-444-53429-3.00002-8
9. McKinney, W. (2018). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython (2nd ed.). O'Reilly Media.
10. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 40(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532
11. Spiegel, M. R., Schiller, J., & Srinivasan, R. A. (2013). Schaum's outline of statistics (4th ed.). McGraw-Hill Education.