Student Performance Analysis for Australian Schools

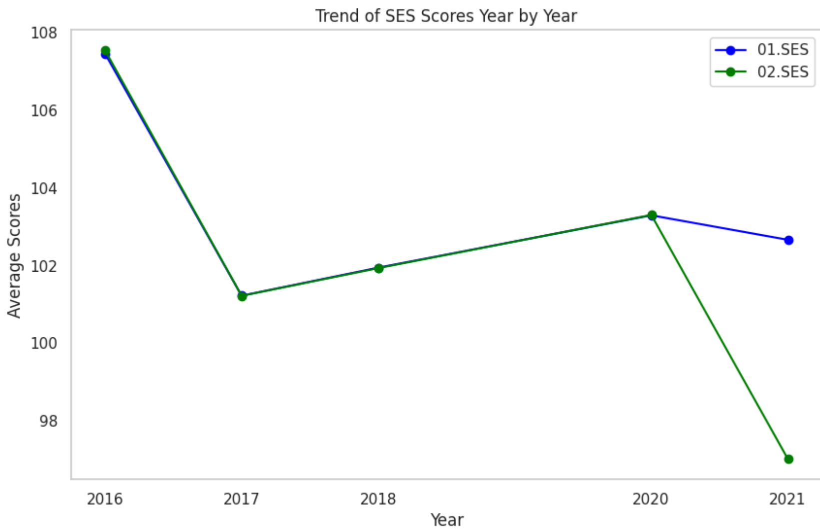
Prepared By: Srijana Bhusal

Executive Summary

This report presents a predictive model designed to identify Year 3 students at risk of underperforming in writing. Using logistic regression and clustering analyses, the model enables early intervention strategies by leveraging students' literacy and numeracy correlations. Key findings indicate that enhancing reading skills can significantly improve writing performance while integrating numeracy exercises supports overall cognitive development. Stakeholders, including students, schools, and consulting firms, stand to benefit through tailored interventions, optimized resources, and valuable insights. Recommendations include expanding data collection, refining algorithms, and implementing standardized processes to enhance model accuracy and educational outcomes.

Table of Content

Introduction



Need

Over the past five years, there has been a rise in the percentage of students who are in danger of performing below expectations in writing. This calls for an examination of the root causes in order to enable efficient treatments.

Change

Our goal is to reduce the number of at-risk students to pre-pandemic levels, to improve academic performance and engagement.

Solution

1. Classification Model: Use logistic regression to quickly identify kids who are at risk so that focused interventions can be made promptly.
2. Clustering Model: To divide students into performance-based groups and implement individualized teaching tactics, apply K-means clustering.
3. Domain Research Solution: Similar to whole language approach to improve speaking ability(Yarmi, 2019), we can use story web, the use of drama and drawings, and explicit instruction approached to improve writing abilities in children(Hess & Wheldall, 1999)

Context

This project is framed within an analysis that the number of students underperforming and at risk in the Year 3 NAPLAN test is on surge.

Stakeholders

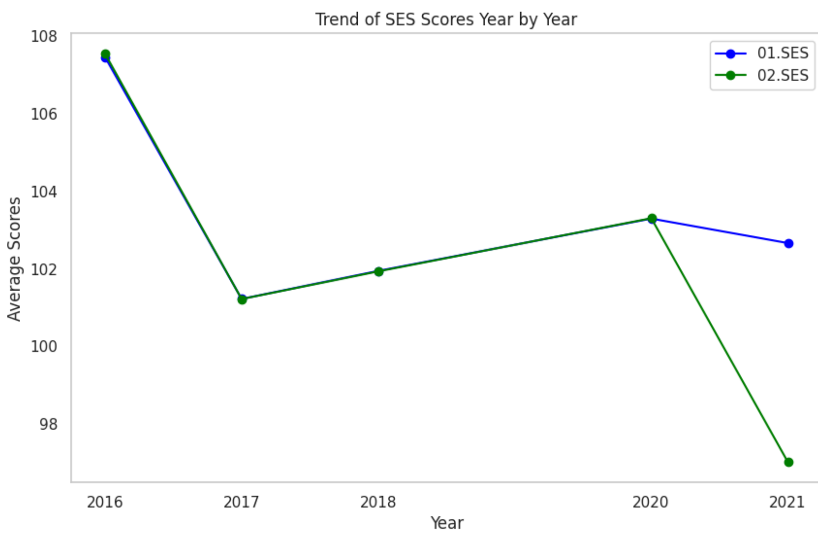
1. Students: They gain from better writing help and instruction as well as data collection through evaluations.
2. Schools: In charge of collecting data and administering tests, schools also put recommendations into practice to improve student outcomes.
3. Consulting Company (Data2Intel): Provides insights and data analysis; project manager is in charge of solution execution and dissemination.

Value

- Students: Success in the classroom and enhanced writing abilities with reduced risk of underperforming on the year 3 NAPLAN test.
- Schools: Having access to efficient techniques for raising student achievement would boost their reputation.
- Consulting Company: Capitalizes on project insights to make a beneficial impact on educational progress.

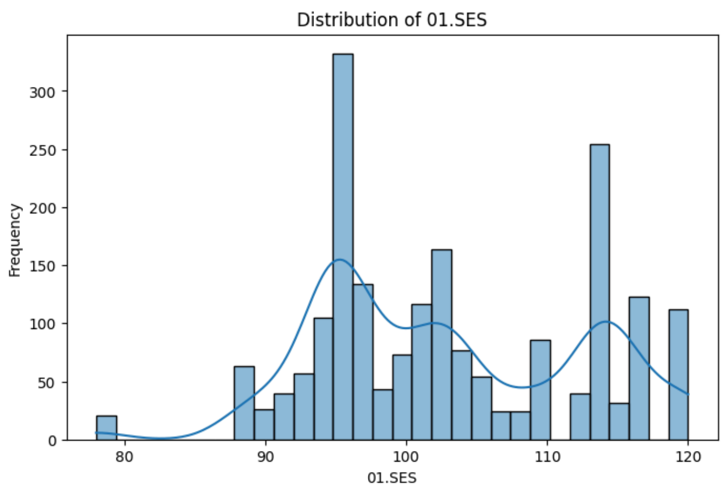
Insights from Exploratory Data Analysis (EDA)

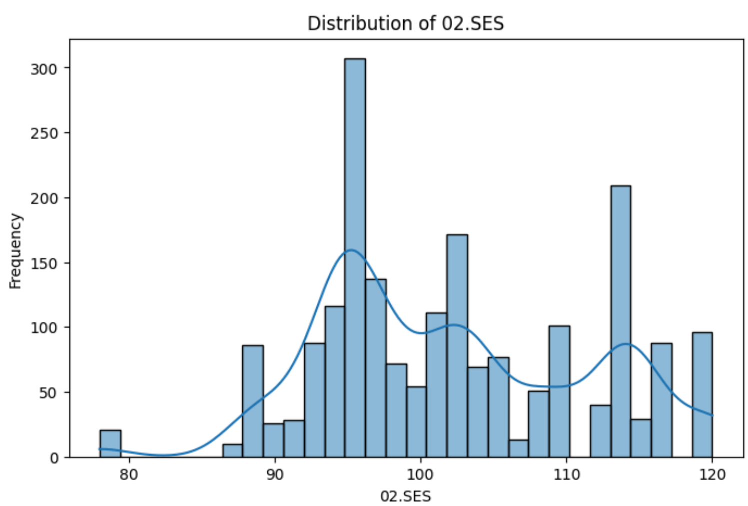
1. SES backgrounds of the students in the dataset in Year 1 and Year 2.



In comparison to national averages, the socioeconomic status (SES) backgrounds of students in the Year 1 and Year 2 datasets show a typically positive standing. With a standard deviation of 9.39 and a mean SES score of 102.94 for Year 1, independent school students' SES is, on average, slightly higher than the national average of 102. With 75% of students scoring below 113 and a minimum score of 78 to a maximum score of 120 for SES, the distribution of scores implies that most students come from very favorable socio-economic circumstances. In Year 2, the range from 78 to 120 is comparable, and the average SES score falls slightly to 102.12 with a standard deviation of 9.15.

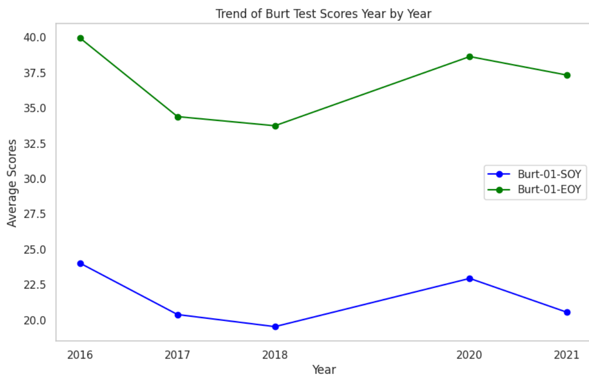
According to the data, the majority of students have socioeconomic backgrounds that are on par with or higher than the national average; this pattern has persisted for both years. Strong favorable correlations between students' socioeconomic status (SES) and their parents' educational backgrounds and professional jobs are found using correlation analysis. Higher SES is associated with families that have achieved more in school, as evidenced by the strong association (0.276) found between the proportion of parents in professional occupations and those with qualifications higher than a diploma (0.181). Conversely, a slight negative association exists between SES and characteristics including the number of siblings and sibling order, showing that bigger family sizes may correlate with lower SES. Kindergarten age and SES have a somewhat positive connection (0.106),

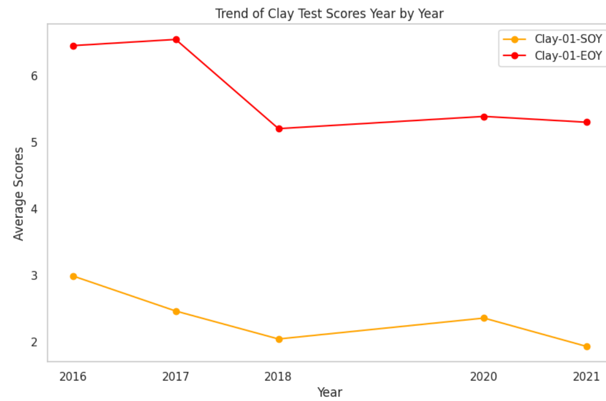


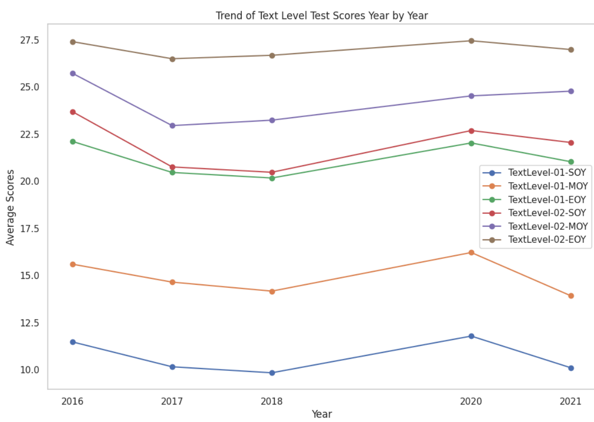


2. Students' reading skills, for example, Burt Reading Scores, at the start and end of Year 1 and at the start and end of Year 2.

Reading skills are determined by the Burt word test, Clay test, and TextLevel test.



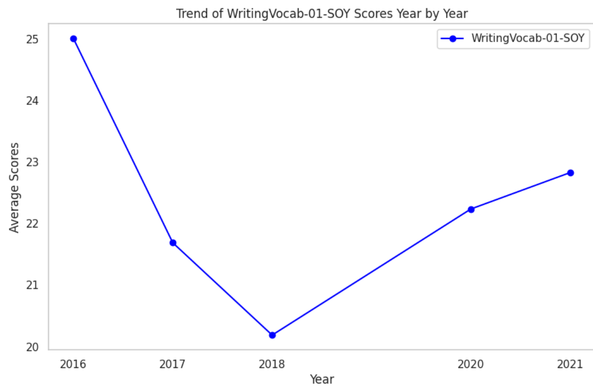


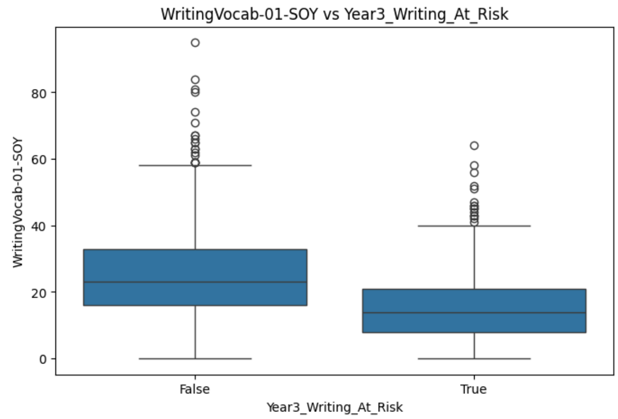


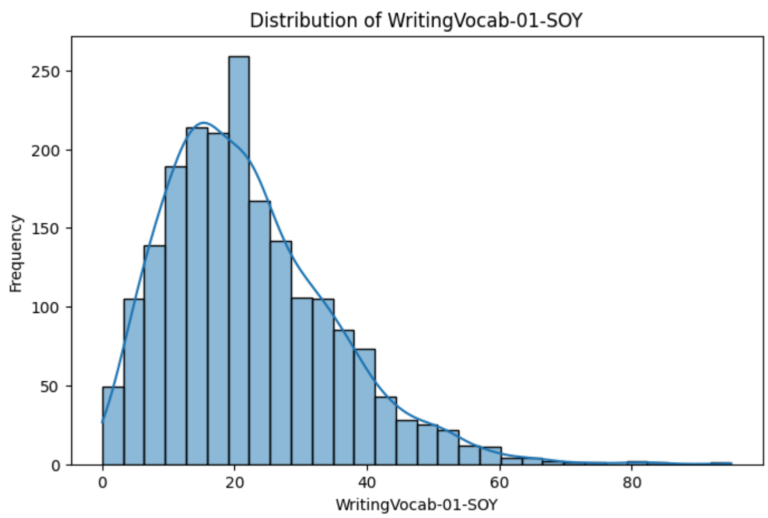
As seen by their results on the Burt Reading Scores, Clay Scores, and Text Level tests, students' reading abilities significantly improved over Years 1 and 2.

The Burt Reading Scores had the most improvement, with the mean score rising from 21.38 at the beginning of Year 1 to 36.57 at the conclusion, demonstrating the efficacy of reading interventions. Clay Scores improved modestly, rising from a mean of 2.14 at the start of Year 1 to 5.31 at the end, suggesting that while progress was made, many students still faced challenges. With a mean of 10.70 at the beginning of Year 1, 21.13 by the end of Year 1, and 26.99 by the end of Year 2, Text Level Scores showed a clear rising trend. Text Level Scores revealed a substantial upward trend, increasing from a mean of 10.70 at the start of Year 1 to 21.13 by the end of Year 1, and further to 26.99 by the end of Year 2. All things considered, the data shows that specific teaching tactics have improved students' reading skills, and the Burt Reading Scores are a particularly useful tool for tracking their progress.

3. Students' writing skills at the start of Year 1, WritingVocab-01-SOY and relationship between this and Year3_Writing_At_Risk.





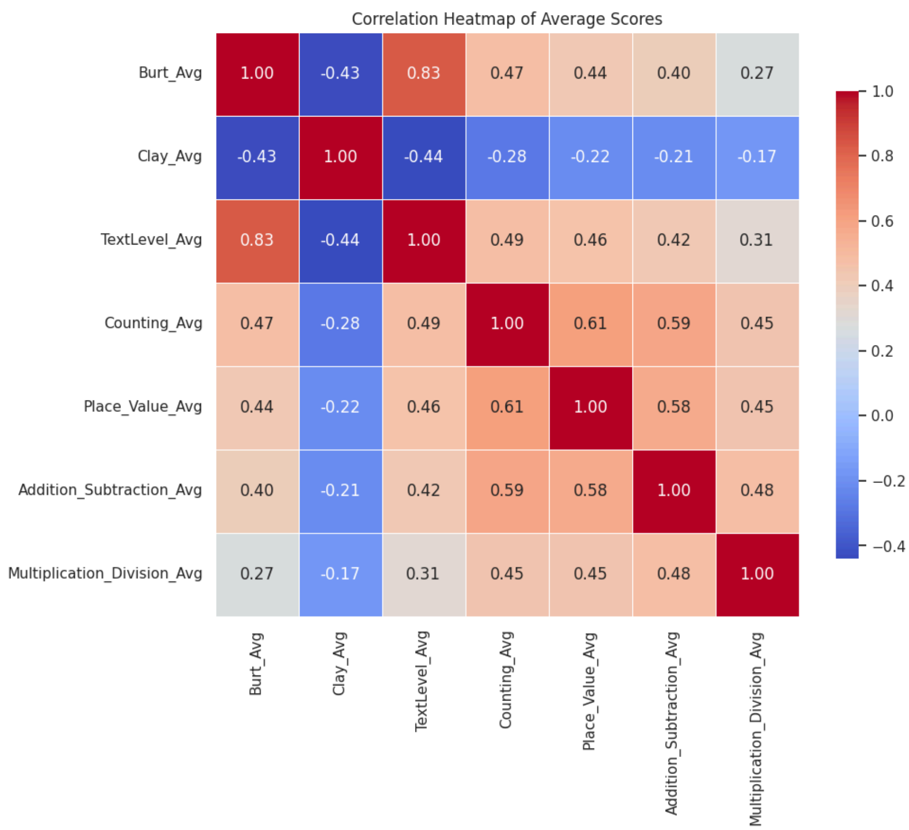


The summary data for WritingVocab-01-SOY, which measures the students' writing skills at the start of Year 1, suggest a mean score of 22.02 with a standard deviation of 12.68. The scores, which vary from 0 to 95, indicate that students' writing vocabulary skills vary greatly from one another.

There is a noticeable distinction between students who are classified as at risk (True) and those who are not at risk (False) when we look at the association between WritingVocab-01-SOY and Year3_Writing_At_Risk. Students who are at risk have a mean score of 15.78, which is significantly lower than the 25.13 score of students who are not in danger. This difference suggests that children who are recognized as being at risk of underperforming in Year 3 NAPLAN writing examinations typically have inferior writing vocabulary skills at the beginning of Year 1. This difference is supported by the lower median (14.0 for at-risk versus 23.0 for not at-risk).

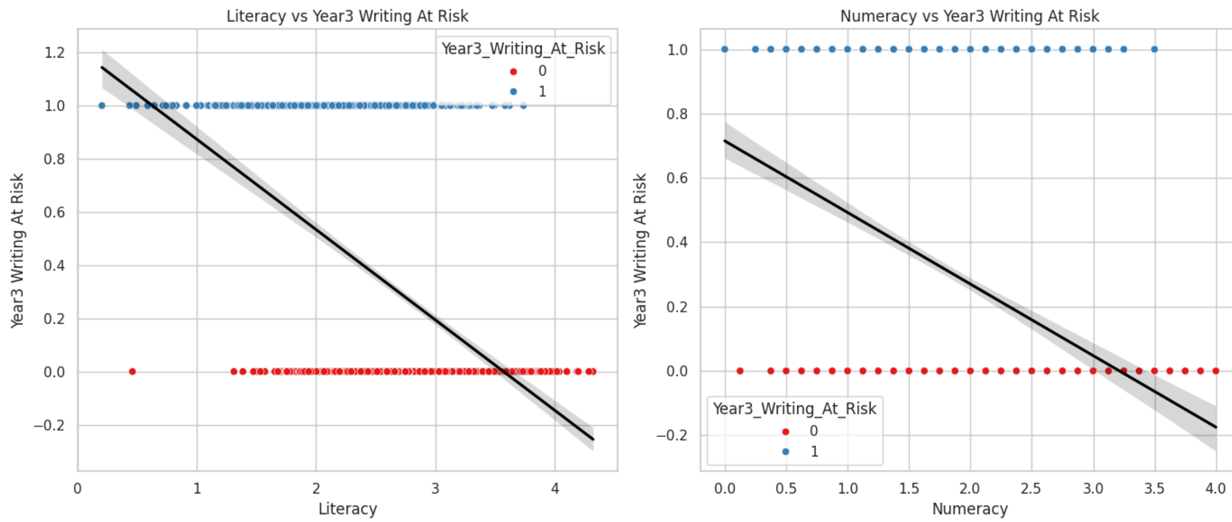
In conclusion, the WritingVocab-01-SOY scores indicate that there is a considerable variation in students' writing abilities and that students who have a lesser vocabulary are more likely to be labeled as potentially underperforming writers by the third year of school. This correlation emphasizes how crucial early interventions are in supporting at-risk students and helping them expand their writing repertoire.

4. Relationship between students' literacy skills and numeracy skills along with relationships between these and their Year3_Writing_At_Risk.



Significant relationships between students' literacy and numeracy skills are shown by the analysis. Burt_Avg and TextLevel_Avg show a strong positive association (0.83), indicating that students who score highly on the Burt test also tend to do well on Text Level assessments. On the other hand, a moderately negative correlation (-0.43) indicates a relationship between higher Burt scores and lower Clay scores for Burt and Clay, respectively. In terms of numeracy, Place_Value_Avg and Counting_Avg correlate 0.61, suggesting that place value proficiency improves counting abilities, while Counting_Avg and TextLevel_Avg have a moderately positive correlation (0.49), suggesting that improved counting skills are related to improved reading comprehension.

Overall, the substantial association between Burt_Avg and TextLevel_Avg indicates that literacy and numeracy skills are tightly related, but the relationships between numeracy and literacy skills are less clear, indicating that in some settings literacy and numeracy may grow separately.



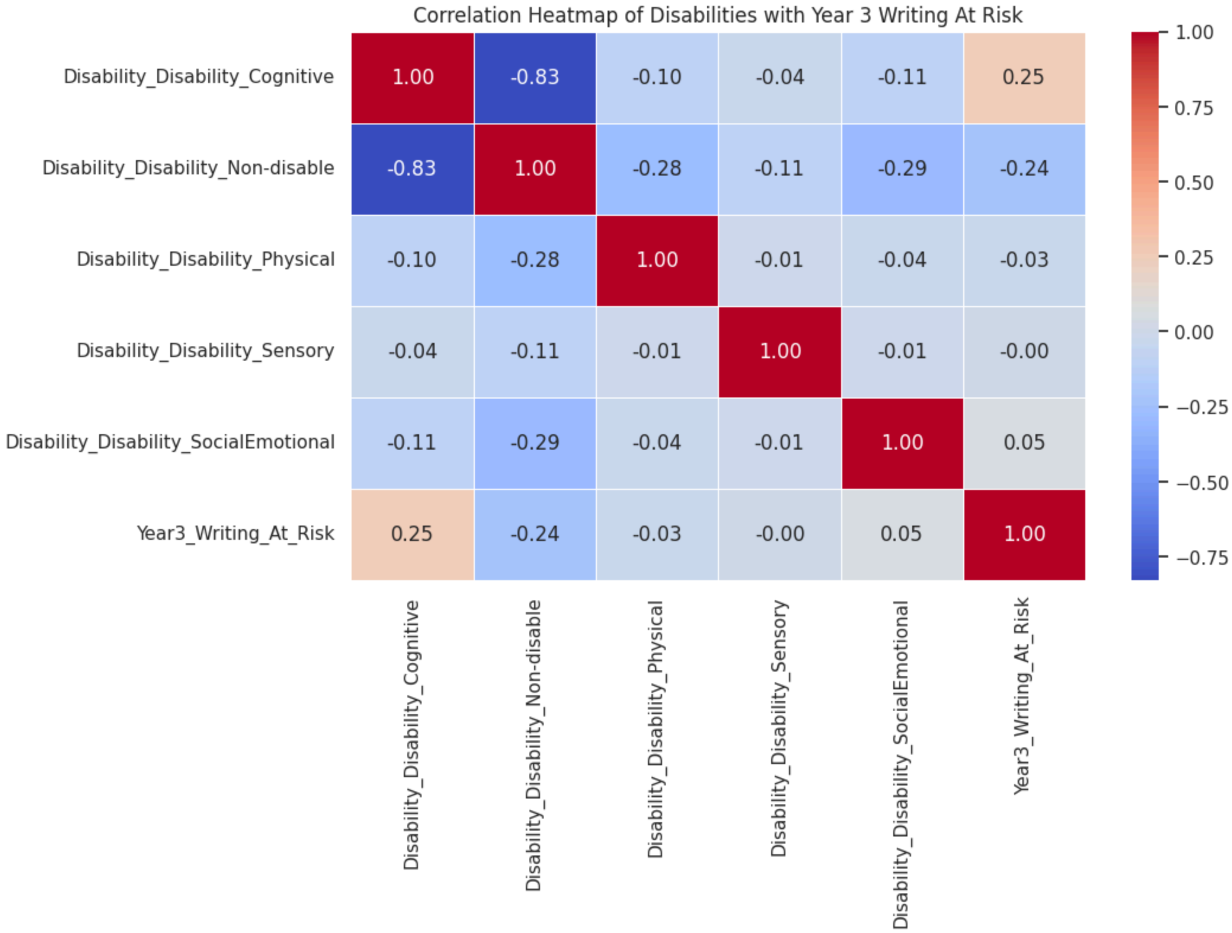
The graphs illustrate the connections between Year 3 Writing At Risk (binary outcome: 0 or 1), Literacy (left plot), and Numeracy (right plot).

Writing in Year 3 at Risk vs. Literacy has a poor indication. The black regression line's downward trend illustrates how the likelihood of being at risk in Year 3 writing diminishes with increasing literacy.

Numeracy vs. Year 3 Writing At Risk: There is also an unfavorable association. The regression line's decreasing slope indicates that when numeracy scores rise, there is a corresponding decrease in the probability of being at risk for Year 3 writing.

Being at risk in Year 3 writing is inversely correlated with both literacy and numeracy skills; greater scores in these domains lower the risk.

5. Students' disability conditions and its relationships with Year3_Writing_At_Risk.



Disability_Cognitive and Year 3 Writing At Risk have a somewhat positive connection (0.25). This shows that Year 3 writing is a riskier subject for students with cognitive difficulties. Students without disabilities are less likely to be at risk, as seen by the negative correlation (-0.24) between Disability_Non-disable and Year 3 Writing At Risk. Disability_SocialEmotional, Disability_Physical, and Disability_Sensory all had poor correlations (-0.03, -0.00, and 0.05, respectively) with Year 3 Writing At Risk, indicating that there is little to no direct association between these conditions and writing risk in Year 3. In conclusion, there is no significant association or a weaker correlation between different forms of disability and Year 3 Writing At Risk compared to cognitive disorders.

6. Insights that might inform early interventions to improve students' writing skills.

The relationships that have been identified offer insights into various early interventions aimed at enhancing students' writing abilities. Burt_Avg and TextLevel_Avg have a substantial positive association, suggesting that improving reading proficiency through focused interventions may enhance writing skills. Furthermore, a moderately positive association has been found between

TextLevel_Avg and Counting_Avg, suggesting that the improvement of numeracy skills may have a good effect on literacy. Additionally, the negative association between Burt_Avg and Clay_Avg shows that even proficient readers may find it difficult to master some writing skills, highlighting the need for specialized education in grammar, spelling, and writing structure. All things considered, these observations highlight the significance of a comprehensive strategy that targets literacy and numeracy abilities to promote improved writing results.

Proposed Machine Learning Solution

In light of our findings, we advise using the logistic regression model to identify students who may perform poorly in writing. By correctly classifying 235 kids as not in danger and correctly identifying 62 students as at risk, the program has proven its competence. The model's recall rate of only 47% suggests that, despite its overall accuracy of 66%, it is still not very good at identifying kids who are in danger. This implies that a sizable portion of at-risk students—71 in total—were mistakenly labeled as not at risk, highlighting the necessity for ongoing model validation and improvement.

Numerous benefits provided by logistic regression are in line with our goals. Because of its ease of use and interpretability, stakeholders can better grasp the probabilistic nature of forecasts and learn about the variables that may be affecting a student's risk of performing below expectations. Furthermore, the model works well when data can be split linearly and is resistant to overfitting, particularly when working with huge datasets. But it's important to recognize its limitations: The log chances of the dependent variable and the independent variables are assumed to have a linear relationship in logistic regression. This may result in oversimplification when dealing with intricate, nonlinear variable interactions. As we proceed, it will be crucial to keep an eye on the model's performance and investigate possible improvements to strengthen its capacity to correctly identify students who are at risk while addressing its shortcomings.

Recommendations and Conclusions

In educational contexts, the use of machine learning models—especially the logistic regression model—can benefit from a variety of business applications. This involves identifying students who are likely to perform poorly in writing early on, allowing schools to properly devote resources and customize interventions. Furthermore, by grouping kids with comparable characteristics for focused support, the clustering analysis might assist schools in creating individualized learning plans.

For all parties involved, the implementation of these predictive models has substantial advantages. Early detection of at-risk students can result in prompt treatments, which will eventually enhance their academic performance and writing abilities. Schools can improve overall educational outcomes, optimize resource allocation, and improve teaching practices by utilizing these findings, which will boost their efficacy and reputation. By using data analysis to obtain insights, consulting firms can enhance their value proposition in the market by improving their educational services and making money.

The integration of machine learning technologies into educational processes can revolutionize decision-making by delivering data-driven insights that influence interventions and instructional approaches. Because schools may now make educated decisions based on predictive analytics rather than depending just on conventional evaluation techniques, this can result in more productive and efficient business processes. Together with better student performance, increased accountability and openness in educational procedures are also possible effects.

It is suggested that more varied data be regularly collected and included to improve model accuracy in order to optimize the efficacy of machine learning solutions(Tiwari, 2023). The dataset will remain strong and dependable if data-gathering procedures are standardized, which will aid in the removal of incorrect results. Furthermore, experimenting with different methods and adjusting hyperparameters could enhance model performance even further. Conducting routine evaluations of intervention results will yield insightful commentary that will facilitate continuous modifications and enhancements of instructional approaches.

References:

Tiwari, R. (2023). The integration of AI and machine learning in education and its potential to personalize and improve student learning experiences. *INTERANTIONAL JOURNAL of SCIENTIFIC RESEARCH in ENGINEERING and MANAGEMENT*, *07*(02). https://doi.org/10.55041/ijsrem17645

Yarmi, G. (2019). Whole-Language Approach: Improve the Speaking Ability at Early years School Level. *JPUD - Jurnal Pendidikan Usia Dini*, *13*(1), 15–28. https://doi.org/10.21009/10.21009/jpud.131.02

Hess, M., & Wheldall, K. (1999). Strategies for improving the written expression of primary children with poor writing skills. *Australian Journal of Learning Disabilities*, *4*(4), 14–20. https://doi.org/10.1080/19404159909546605