ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054

Evelance - Evaluating Neural Network Predictions of Student Academic Performance using Synthetic Data

Atharv Narang
C.L. Gupta World School

Abstract
There is a link between the academic performance of students, their study habits, attendance, sleep, and computer usage; however, forecasting examinations based on these traits is difficult. While machine learning algorithms are capable of finding statistical correlations between student characteristics, it is not enough for the prediction to be reliable for other student profiles or incomplete inputs. In this experiment, the author has studied if a neural network can predict the results of the exam better than a regular regression algorithm and if there are any changes in the performance of the model in controlled experimental conditions.

An experiment with a neural network written in PyTorch, consisting of three hidden layers (32, 64, and 128 neurons) has been carried out using 10,000 artificial student records that contained six academic and behavioral characteristics. The experiment included ten tests related to comparing model performances, repeated validation, contributions of different features, errors in subgroups, characteristics of the dataset used, shortcuts in predictions, training-set size, sensitivity to missing data, and controlled changes to model inputs.

The neural network reached the result of mean absolute error of 5.7784 and R^2 of 0.7358, which outperformed the results of the linear regression (mean absolute error – 6.9173) and random forest regression (mean absolute error – 5.8657). In five repeated tests, it gave the mean MAE of 5.7206 ± 0.1821. However, it has proven to be much worse at predicting the results of low-scoring records (mean absolute error 19.7964) than those of the high-scoring records (mean absolute error 4.9707). At 40% randomly missing input values, mean MAE increased by approximately 48.4%.

I. INTRODUCTION

What if one of the most significant barriers to a student's performance academically, which could possibly have been declining rapidly, was actually in their pockets? According to the results from OECD PISA 2022 survey, almost 30% of students in OECD member countries indicated that they were getting distracted in math lessons by their digital devices, which resulted in them having 15 points less than those students who were rarely getting distracted from math classes, even taking into consideration their socioeconomic status (Cormann). Digital distractions is only one of the multiple factors that impact academic performance of a student, but other important factors include sleeping habits, proper nutrition and study environment. The purpose of the research is to find out if machine learning algorithms are capable of predicting the factors that affect academic performance of a student such as sleeping habits, study environment, and nutrition.

II. LITERATURE REVIEW

Sleep and Academic Performance

Adolescents require approximately 9–9.5 hours of sleep per night. However, approximately 80% of adolescents fall short of the recommended amount of sleep, and around 28% of students report falling asleep at school because of fatigue. Insufficient sleep can negatively affect attention, executive functioning, and overall learning ability.

Both acute and chronic sleep deprivation can negatively affect academic performance. Even a single night of severe sleep restriction can reduce attention and reaction time, while repeated sleep deprivation can impair memory consolidation and higher-order reasoning. Approximately 24 hours without sleep can produce cognitive impairment comparable to roughly a 0.10% blood alcohol concentration.(Kim , Lee et al.) Chronic sleep problems are also associated with anxiety, depression, and weakened immune functioning.

Attendance and Academic Performance

Regular attendance is an important predictor of both short-term and long-term educational success. Missing school interrupts learning sequences, decreases exposure to instructional material, and allows learning gaps to accumulate over time.

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054

Chronic absence is generally defined as missing at least 10% of school days, equivalent to approximately 18 or more days during an academic year.(InterDynamics) Even missing two days per month can place students at greater risk of falling behind. The effects of absenteeism can also extend beyond the individual student because teachers may need to reteach material for students returning after absences, potentially disrupting classroom pacing. Missing approximately 10% of the school year substantially decreases the probability of achieving grade-level outcomes on time and has been associated with an estimated 25% reduction in the likelihood of college enrollment.(Shankar,Dooley et al.)

Digital usage and Academic Performance

Digital behavior is another factor associated with student academic performance. Internet use, gaming, social media, and digital multitasking can fragment attention, displace time that could otherwise be spent studying or completing homework, and interfere with healthy sleep patterns. Heavy nighttime device usage may shorten sleep duration and delay circadian timing.

Problematic internet and gaming behaviors have also been associated in observational research with lower grades and increased social and emotional difficulties.

Health and Student Wellness

Student health can indirectly influence academic outcomes through its effects on attendance and classroom engagement. On-campus health services, asthma-management programs, and mental-health support can reduce health-related absences and improve sustained attendance.

Integrated wellness programs that address multiple aspects of student health and behavior tend to generate stronger attendance and learning improvements than isolated interventions. This suggests that academic performance may be better understood by examining several lifestyle, behavioral, and environmental variables together rather than treating each factor independently.

Student Engagement

Repeated absence, insufficient sleep, and unmet health or social needs can contribute to student disengagement. Approximately 40–60% of high-school students report chronic disengagement in some surveys.(Crotty)Potential strategies for improving student engagement include mentoring and outreach, trauma-informed approaches, and stronger relationships between students and school staff.

Overall Findings

Academic performance is influenced by a network of interconnected lifestyle, behavioral, health, and environmental factors. Sleep duration, attendance, digital usage, nutrition, health, and student engagement may influence academic outcomes both directly and indirectly.

The interaction among these variables provides a basis for using machine-learning models to analyze multiple student characteristics simultaneously. Such models can be used to investigate patterns between behavioral variables and academic outcomes and potentially identify which factors are most strongly associated with differences in student performance.

III. METHODOLOGY

The dataset consists of synthetic student academic records from high school students, and includes factors like attendance, study hours, previous exams scores, hours of sleep, social media usage of students, and the students' performances in their exams. It contains 10,000 samples and we got the dataset from https://www.kaggle.com/datasets/shambhurajejagadale/student-performance-prediction-dataset?resource=download.

The model is a neural network model which uses PyTorch Library and includes several different layers. The first hidden layer contains a linear layer followed by a batch normalization layer, then a rectified linear unit, and finally a dropout layer and contains 32 neurons. The second hidden layer also follows the same sequence and contains 64 neurons.This order is followed by a third and the last hidden layer followed by an output layer which contains a linear layer which takes 128 outputs from the last hidden layer and produces a single output which is the exam score. evaluate the model's robustness to incomplete information, random input values in the testing dataset were masked at five missingness levels: 0%, 10%, 20%, 30%, and 40%. Missing values were replaced using medians calculated from the training data. The model was evaluated without retraining, and five independently generated missingness masks were used at each level to assess variation in predictive performance. MAE, RMSE, and R^2 were calculated for each condition.

The Dataset was divided into training and testing subsets with a 80:20 ratio which was of 8,000 training records and 2,000 testing records.Input features were standardized using (standard scaler) preventing information from the testing subset to influence preprocessing. The model was trained using the Adam optimizer with a learning rate of 0.01 and MSE(MEAN SQUARED ERROR) as the loss function . The experiment configuration used 500 training epochs.The performance was evaluated using MAE(Mean

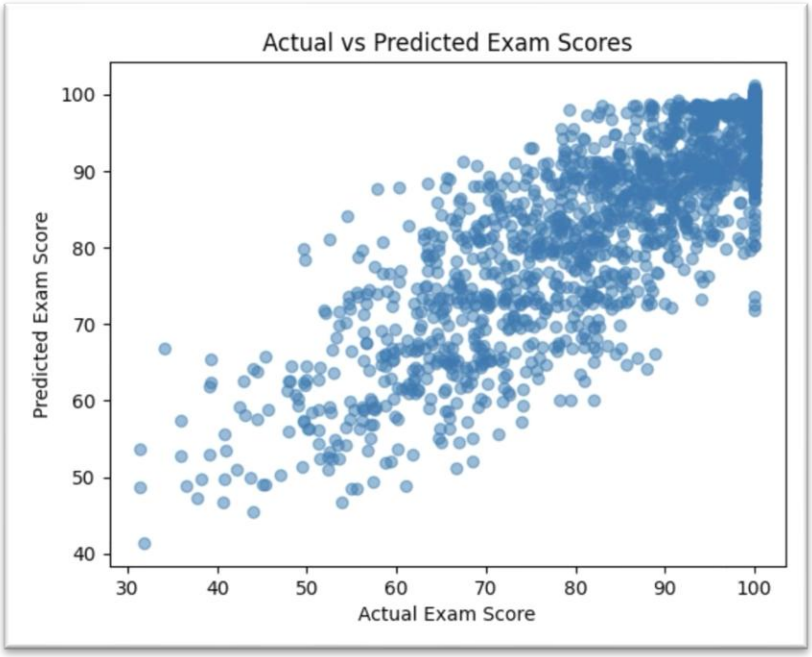
ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054

Absolute Error) , RMSE(Root Mean Squared Error) and R^2(coefficient of determination).

IV. RESULTS AND CONCLUSIONS BASELINE

The Scatterplot Illustrates the Comparison Between Actual Student Exam Scores Along the X-Axis and Predicted Scores Generated by The Model along the Y-Axis, Highlighting A Positive Correlation between both Variables.The Dataset Consists of Synthetic Student Academic Records from High School Students, and Includes Factors Like Attendance, Study Hours, Previous Exams Scores, Hours of Sleep, Social media usage of students, and the students' performances in their exams. It contains 10,000 samples and we got the dataset from https://www.kaggle.com/datasets/shambhurajejagadale/student-performance-prediction-dataset?resource=download.

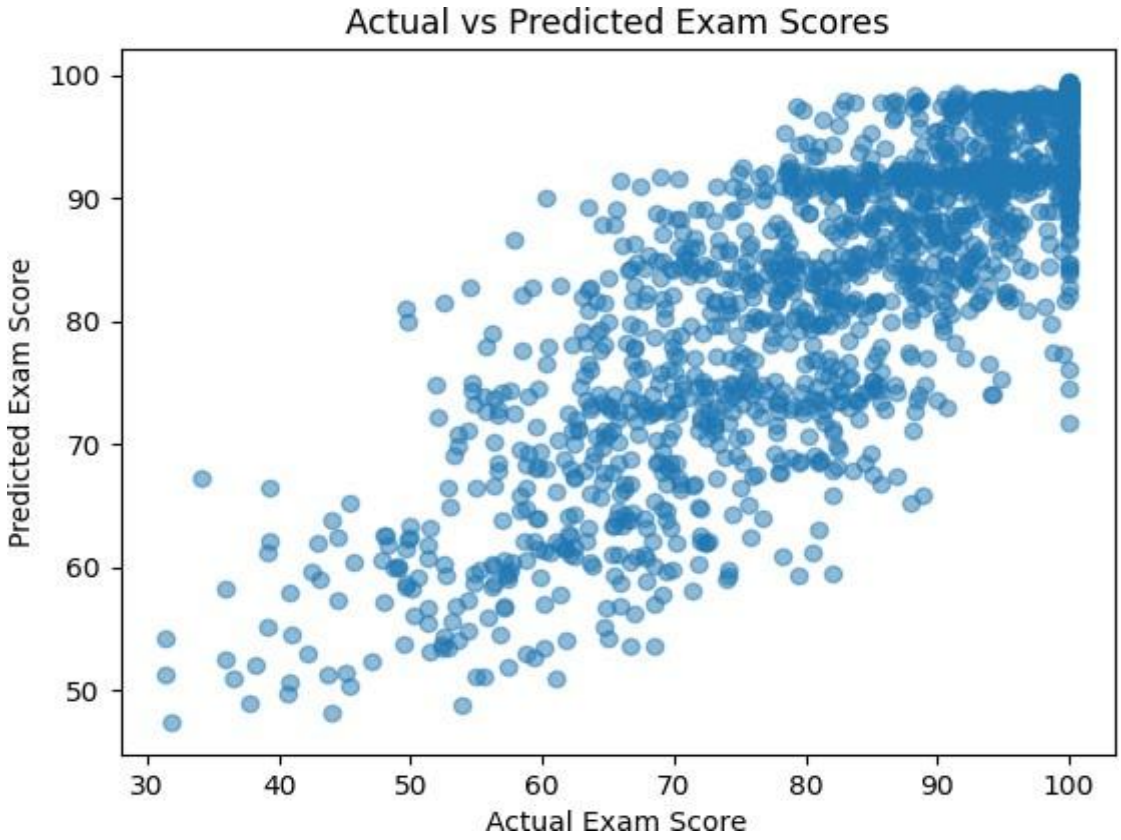
The model is a neural network model which uses PyTorch Library and includes several different layers. The first hidden layer contains a linear layer followed by a batch normalization layer, then a rectified linear unit, and finally a dropout layer and contains 32 neurons. The second hidden layer also follows the same sequence and contains 64 neurons. This order is followed by a third and the last hidden layer followed by an output layer which contains a linear layer which takes 128 outputs from the last hidden layer and produces a single output which is the exam score.The study demonstrates that machine learning methods can be used to model the complex relationship between student academic performance and other behavioural factors such sleep or nutrition.This can be very useful for students to understand their shortcomings.However, we must note that the model was trained on a synthetic dataset meaning that the performance generated by this model cannot be directly to applied to the student population immediately.



Experiment A : Comparative Model Performance

The neural network model of Evelance was compared with a linear regression model and a random first regression model using Mean Absolute Error(MAE) , root mean squared error (RMSE) and coefficient of determination (R^2). The neural network model got an MAE of 5.7784 , RMSE of 7.7913 and R^2 score of 0.7358 and it outperformed both random first regression (MAE = 5.8657, RMSE = 8.1195, R^2 = 0.7130) and the linear regression model(MAE= 6.9173, RMSE = 8.7932, R^2 = 0.6634).

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054



Experiment B : Repeated Validation

This experiment was conducted to check the consistency of the performance of the model but this time across five random seeds which were 42, 52, 62, 72, 82. The evaluations showed that the neural network model had a mean MAE of 5.7206± 0.1821 , mean RMSE of 7.7951 ± 0.1444 and mean R^2 of 0.7277 ± 0.0071.Random forest regression model had mean MAE of 5.9369 ± 0.1455 and mean R^2 of 0.6991 ± 0.0099 whereas linear regression model achieved a mean MAE of 6.9639 ± 0.1133 and mean R^2 of 0.6550 ± 0.0059.The neural network model had better metrics than the comparison models in all the five trials.This suggests that its advantage isn't confined to a single train test split.

Experiment C : Feature Ablation

This experiment was done to determine that if we remove individual features then does it affect the prediction accuracy of the model.The results show that when no features were removed then the MAE(Mean Absolute Error) was 6.3508 and r^2 was 0.7092.The metrics changed when we removed study hours and the MAE was 9.6541 and the r^2 score was 0.3537. We also saw considerable changes when we removed other variables individually.

Removed feature	MAE	R^2
None (baseline)	6.3508	0.7092
Study hours	9.6541	0.3537
Attendance	6.8182	0.6614
Sleep hours	6.1834	0.7070
Internet usage	6.2754	0.7033
Assignments completed	8.1040	0.5332
Previous score	7.8108	0.5907

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054

Experiment D : Feature Group Comparison

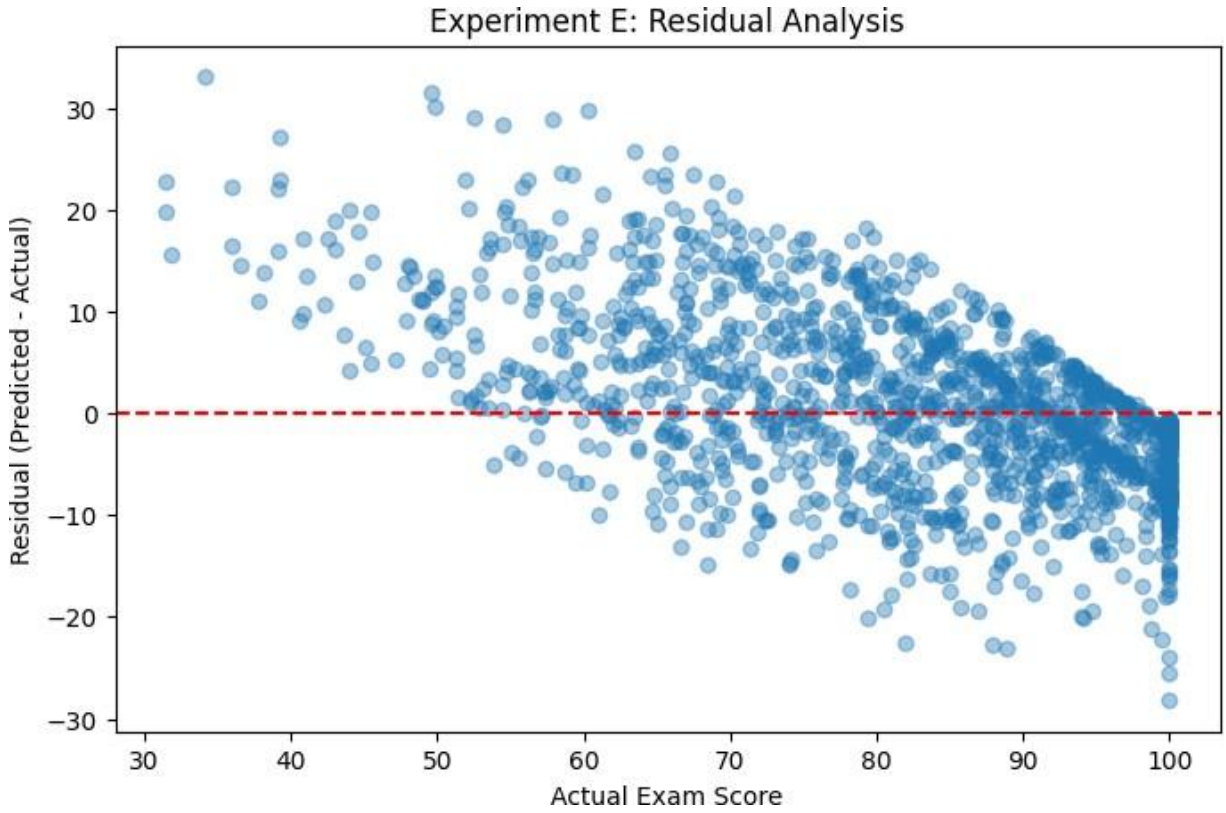
The motive to conduct this experiment was to compare that do the academic variable only model would outperform the full model and the lifestyle variable only model. The results proved that the full feature model was the strongest whereas the lifestyle variable only model had very little variation in the examination scores and the model with academic variables but no previous score had worse results than the model with previous scores as it gave additional information.

Feature group	MAE	R^2
All six features	5.7784	0.7358
Academic only	6.2938	0.6758
Lifestyle only	12.3269	0.0285
Academic without previous score	7.5546	0.5614

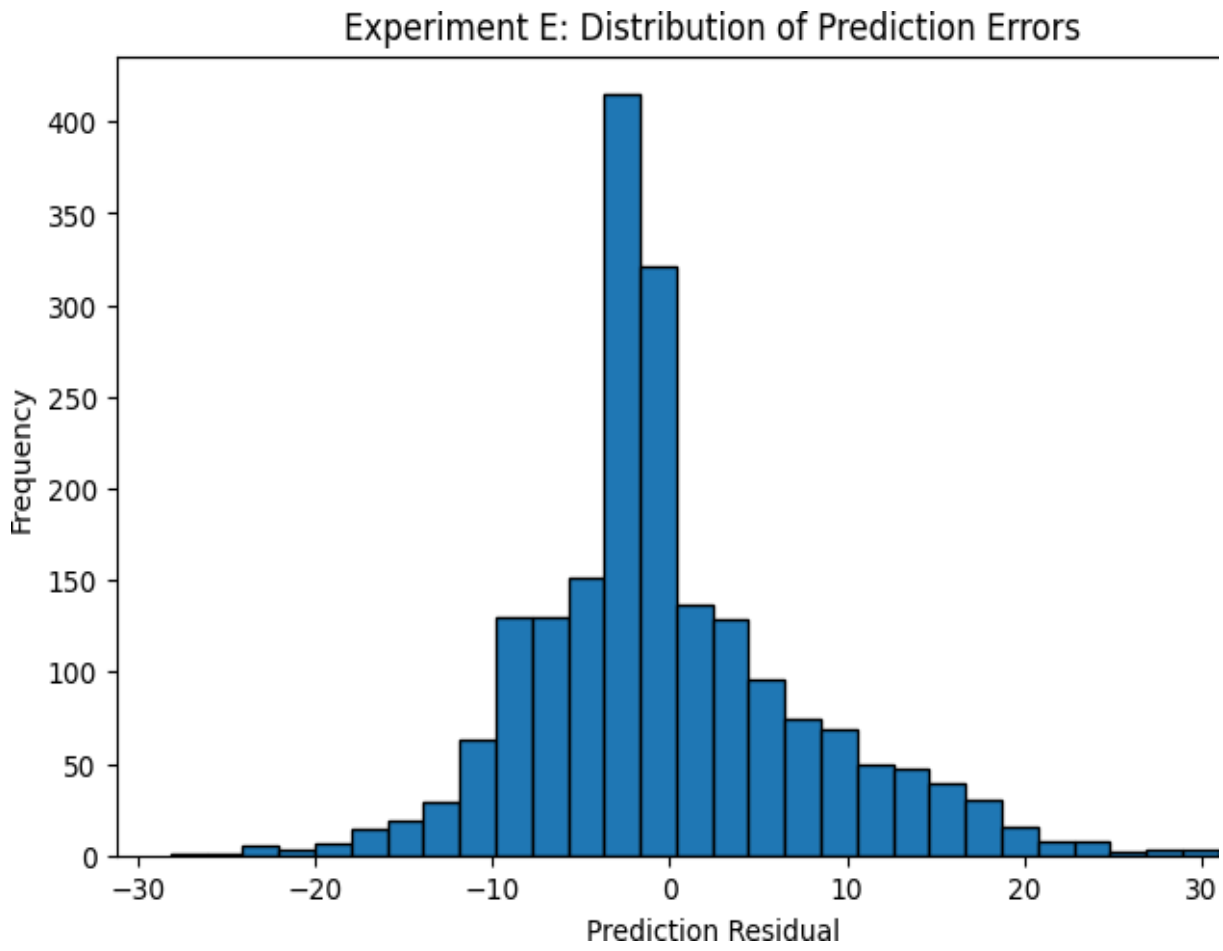
Experiment E : Prediction Error Analysis

The objective of the experiment was to check whether model accuracy varied across examination score groups. The results show that the accuracy was not uniform across the score groups and the low score group had approximately four times the MAE of the high score group. The overall MAE was 5.7784 and the overpredictions were 738 or 36.9% whereas there were 1262 underpredictions or 63.1%.The best performance was from the high scoring group followed by the medium scoring groups and in the end were the low scoring groups.

Score category	Records	MAE	Mean residual
Low (≤40)	13	19.7964	+19.7964
Medium (>40–70)	332	9.2558	+7.2639
High (>70)	1,655	4.9707	−1.7480



ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054



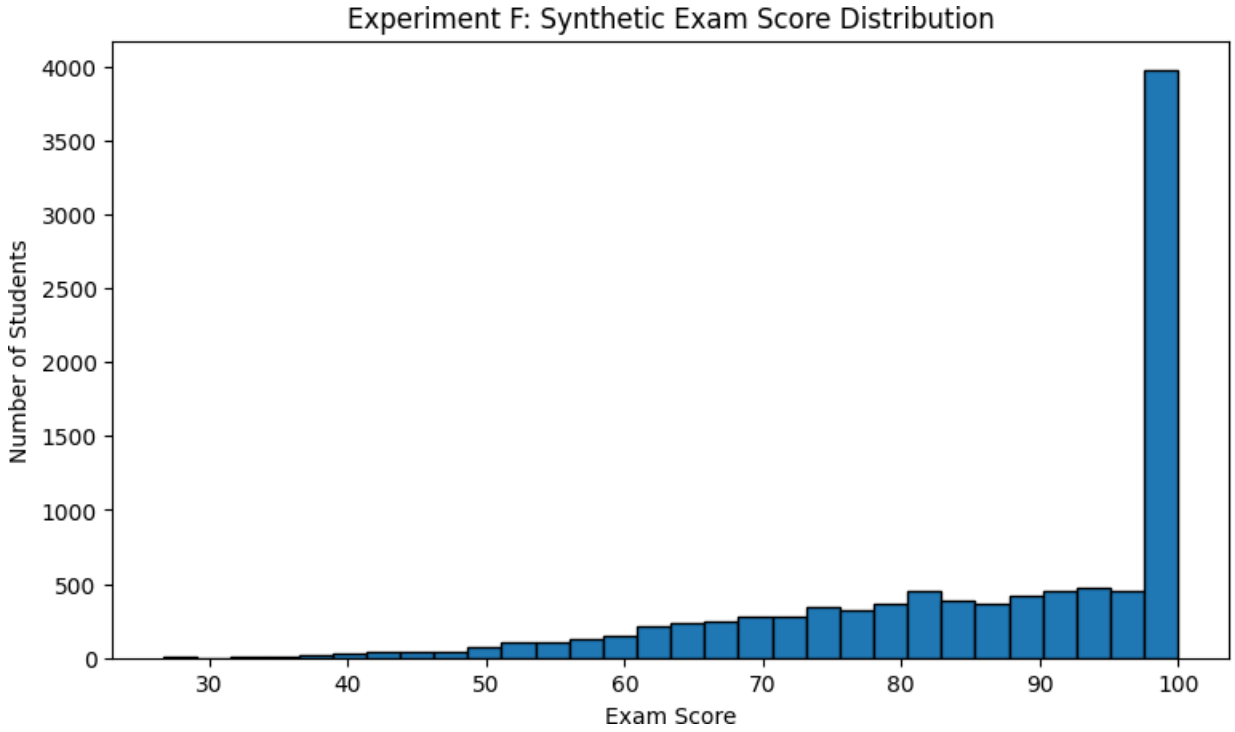
Experiment F : Dataset Audit

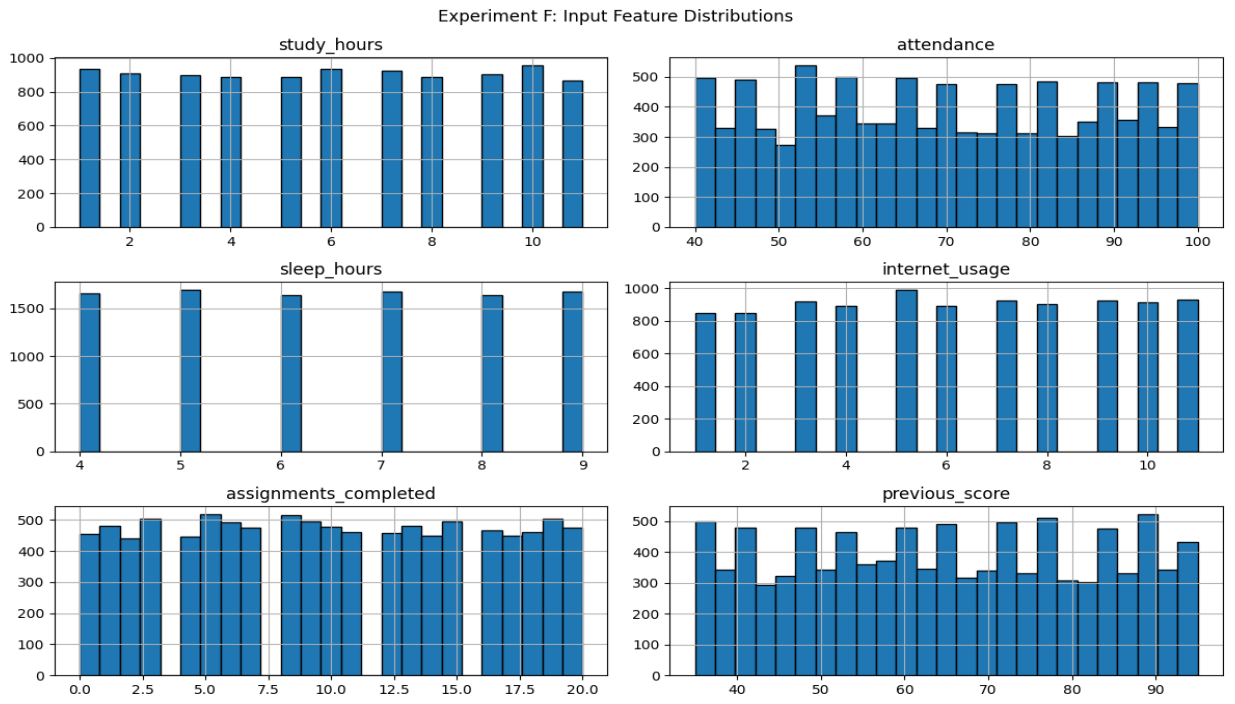
The aim of the experiment was to examine the distribution and statistical characteristics of the synthetic dataset. The results were unexpected as the exam score were concentrated towards the upper end of the scale and more 33% had a score of 100. On the other end low scores were rare and this imbalance may help explain the subgroup performance differences observed in Experiment E , But we must understand that correlation measures association instead of causation

Measurement	**Result**
Total records	10,000
Mean examination score	86.704
Median examination score	92.12
Standard deviation	15.058
Scores ≤40	45 (0.45%)
Scores >70	8,356 (83.56%)
Scores exactly 100	3,520 (35.20%)
Missing values	0
Complete duplicates	0

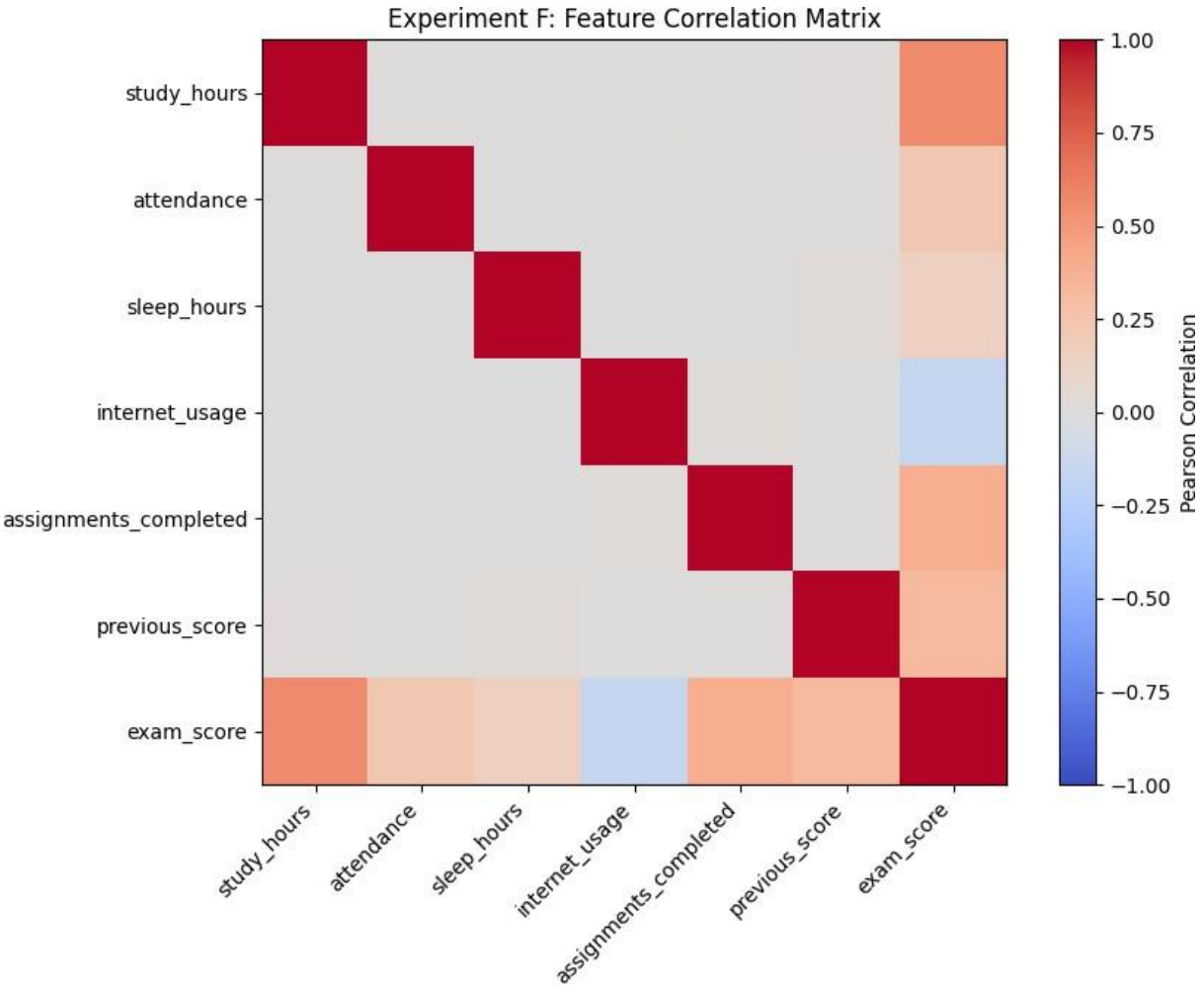
ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054

Feature	Pearson correlation
Study hours	+0.5625
Assignments completed	+0.3876
Previous score	+0.3188
Attendance	+0.2234
Sleep hours	+0.1447
Internet usage	−0.1519





ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054



Experiment G : Leakage and Shortcut Testing

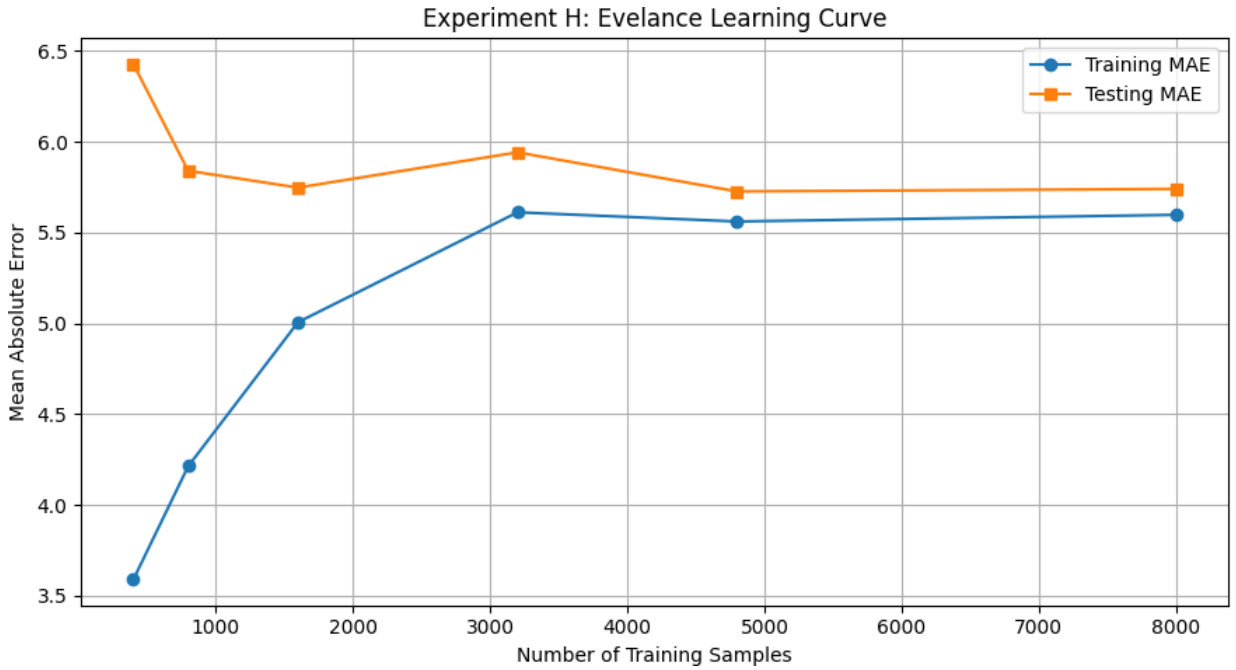
The main objective of this experiment was to investigate whether the model's performance is dependent upon shortcuts or trivial prediction strategies. The results show that the full model had the best performance followed by the model without previous score. Both the shuffled training targets and the training mean baseline had a MAE of 12.9016 and 12.5973 which is more than the double of the full model which was 5.7784.This suggests that the model learnt statistical relationships instead of simply predicting the dataset mean.

Test condition	MAE	R^2
Full model	5.7784	0.7358
Without previous score	7.0538	0.6191
Shuffled training targets	12.9016	−0.0270
Training-mean baseline	12.5973	−0.0004

Experiment H : Learning Curve

The objective of the experiment was to understand the relationship between training set size and prediction accuracy. The results were interesting as the r^2 score generally improved as the training set size increased but the Test MAE did not improve consistently at every step and the performance gain became smaller at larger training sizes but the difference between the training and the testing MAE became narrower as more data became available.

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054



Training records	Test MAE	Test R^2
400	6.4282	0.6610
800	5.8404	0.6829
1,600	5.7477	0.7205
3,200	5.9420	0.7279
4,800	**5.7267**	0.7314
8,000	5.7404	**0.7360**

Experiment I : Missing Data Robustness

The aim of the experiment was to investigate that hos incomplete input information would affect the prediction accuracy. The results show that accuracy was deterring as the more information goes missing. R^2 decreased from 0.7360 to 0.4164 and the experiment used randomly missing values and did not retrain the model.

Missing inputs	Mean MAE	Mean R^2
0%	5.7404	0.7360
10%	6.5054	0.6552
20%	7.1811	0.5798
30%	7.8175	0.5040
40%	8.5193	0.4164

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054



Experiment J : Counterfactual Sensitivity

The aim of this experiment was to examine how modifying one input variable changes the model's predicted exam score.The model generally predicted higher scores when study hours , attendance , sleep duration and completed assignments increased whereas increasing internet usage generally reduced predicted scores. Modification of variables was done individually and the other variables remained constant at that time.

Feature modification	Eligible records	Mean predicted change
Study hours +1	1,830	**+2.5183**
Attendance +5 percentage points	1,825	+0.9543
Sleep hours +1	1,677	+1.2001
Internet usage +1	1,814	**−0.7694**
Assignments completed +1	1,903	+0.9071
Previous score +5	1,830	+1.2942

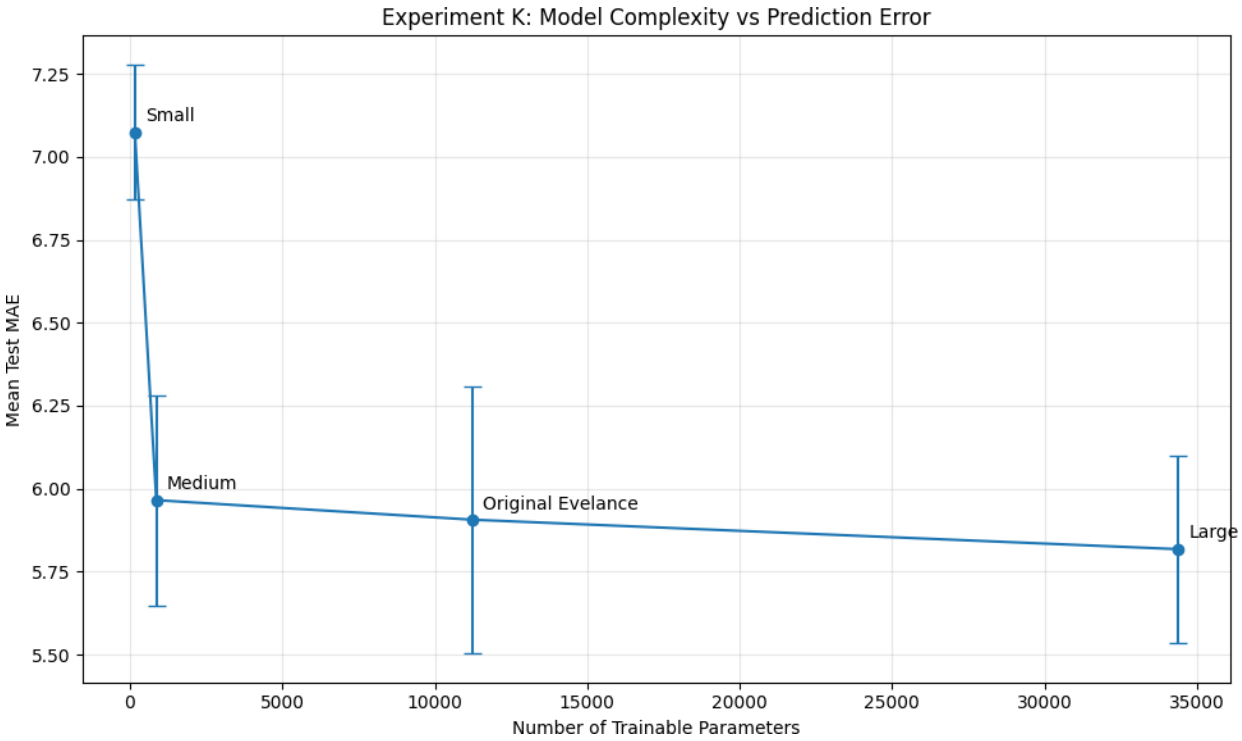
Experiment K : Model Complexity Analysis

The aim of this experiment was to evaluate how the neural network architecture , parameter count and training time influenced the prediction performance. Four neural network structures ranging from 161 to 34,369 parameters were evaluated across 5 random seeds and it produced 20 experimental runs. The large architecture achieved the lowest MAE of 5.8176 while the original Evelance structure achieved the lowest RMSE of 7.8470 and the highest mean r^2 of 0.7240.

Architecture	Hidden layers	Parameter s	Mean test MAE ± SD	Mean RMSE	Mean R^2	Mean training time
Small	1	161	7.0735 ± 0.2031	8.9531	0.6408	128.24 s
Medium	2	865	5.9653 ± 0.3166	8.0301	0.7110	163.81 s

ISSN :2584-2706 DOI: https://doi.org/10.67317/IJMSRT26NOV054

Original Evelance	**3**	**11,233**	5.9064 ± 0.4016	**7.8470**	**0.7240**	229.44 s
Large	4	34,369	**5.8176 ± 0.2816**	7.8640	0.7228	321.29 s



V. DISCUSSION

The findings of this study suggest that the prediction accuracy alone isn't enough as from the evidence above we interpret that the Evelance Neural Network Architecture achieved an MAE of 5.7784 and offered a predictive advantage .Each variable contributed to the prediction accuracy but the academic variable contained more predictive information in the synthetic dataset . We also got to know that the model was sensitive to incomplete information even though it had relatively consistent performance across repeated validation runs.Evelance's most important finding is not simply that a neural network can predict examination scores rather that apparently strong overall accuracy can hide limitations in reliability of the model itself.

REFERENCES

[1]. Cormann, Mathias. "Full Report: Pisa 2022 Results (Volume I) | OECD." *OECD*, 5 Dec. 2023, www.oecd.org/en/publications/pisa-2022-results-volume-i_53f23881-en/full-report.html.

[2]. Kim SJ, Lee YJ, Cho S, Cho I, Lim W, Lim W. Relationship Between Weekend Catch-up Sleep and Poor Performance on Attention Tasks in Korean Adolescents. Arch Pediatr Adolesc Med. 2011;165(9):806–812. doi:10.1001/archpediatrics.2011.128 https://jamanetwork.com/journals/jamapediatrics/fullarticle/1107586.

[3]. InterDynamics. "Is Being Sleep Deprived the Same as Being Drunk?" *InterDynamics*, 2 Apr.2023, www.interdynamics.com/2015/09/01/sleep-deprivation-blood-alcohol/.

[4]. Institution of Educational Sciences. "Chronic Absenteeism." *IES*, 12 Sept. 2022, ies.ed.gov/use-work/supporting-recovery-with-evidence-based-practices/chronic-absenteeism.

[5]. Shankar M, Dooley DG, Cholera R. A Role for Health Systems in Addressing Chronic Absenteeism. JAMA Pediatr. 2024;178(8):743–744. doi:10.1001/jamapediatrics.2024.1743https://jamanetwork.com/journals/jamapediatrics/article-abstract/2819780

[6]. Crotty, James Marshall. "Motivation Matters: 40% of High School Students Chronically Disengaged from School." *Forbes*, 13 Mar. 2013, www.forbes.com/sites/jamesmarshallcrotty/2013/03/13/motivation-matters-40-of-high-school-students-chronically-disengaged-from-school/.