Predicting AI Usage's Perceived Impact on Grades

William White

COLGATE UNIVERSITY

Introduction + Background

The rapid adoption of artificial intelligence (AI) tools raises questions about how AI usage impacts productivity. Students in particular have adopted large language models (LLMs) at a breakneck pace, but oftentimes use it to complete assignments for them instead of leveraging it as a learning tool.

- ➔ Research Question: How does the way students use AI tools contribute to the perceived impact AI has on their academic performance?
- ➔ Hypothesis 1: Students who use AI for learning will believe it has a better impact on their grades than students who use it for assignment completion.
- ➔ Hypothesis 2: Students who use AI for longer periods will believe it has a better impact on their grades.
- ➔ Dataset: ~3000 survey responses from Indian undergraduate students tracking AI daily usage hours, trust in AI tools, and the student AI use cases
- ➔ Method: I will train AI models to predict a student's impact on grades, then I will use factor importance calculations to see how certain variables contribute to predictions and draw conclusions about those variables

Experimental Setup

Data Processing

- ➔ Encoded multi-select and categorical variables
- ➔ Made "consolidated" dataset, removing and combining some variables (ex. combined "project work" and "assignment completion" use cases) (23cols v 42cols)
- ➔ Many alternate data frames to track prediction ability for more/less extreme respondents (i.e. only students with grade impacts between -2 and 2)

Training Models

- ➔ Measured correlation coefficients between variables, which was common among similar research
- ➔ Created a "mean model" function to easily compare model to random guessing, providing a useful performance metric relevant for all data frames

```
df3 - Model RMSE: 1.2718; Model Improvement: 35.86%; Mean RMSE: 1.9829
cdf3 - Model RMSE: 1.3340; Model Improvement: 32.73%; Mean RMSE: 1.9829
```

- ➔ Trained machine learning models (Linear Regression Models, Random Forest Models, and multi-layer FNNs) with all dataframes, comparing their scores

Analysis

- ➔ Compare performance for different data frames
- ➔ Could get coefficients of Linear Regression models
- ➔ Apply SHAP algorithm to Random Forest Models and an FNN
  - ◆ Use charts to determine directionality
  - ◆ Use SHAP importance for degree of impact

Results

Correlation Coefficients

	Year...	...Hours	Awareness...	Trust...	Impact...
Year_of_Study	1.000	0.008	-0.013	0.001	-0.006
Daily_Usage_Hours	0.008	1.000	0.006	-0.012	0.057
Awareness_Level	-0.013	0.006	1.000	0.039	-0.016
Trust_in_AI_Tools	0.001	-0.012	0.039	1.000	-0.047
Impact_on_Grades	-0.006	0.057	-0.016	-0.047	1.000

Correlation Coefficient table for the variables which were initially numeric

For all variables across all data frames, there were no relevant strong linear relationships between variables.

Model Improvement Scores

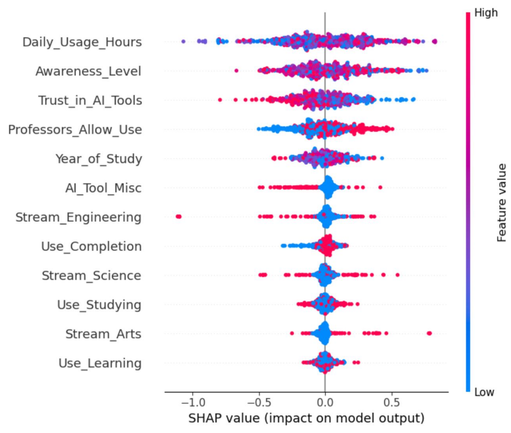
	df	cdf	df3	cdf3	df2	cdf2	df1	cdf1	dfp	cdfp	dfn	cdfn
Linear Regression	-1.31%	-1.01%	1.57%	1.85%	3.19%	3.80%	-0.82%	-0.51%	1.33%	3.58%	-0.47%	0.60%
Random Forest	18.67%	16.04%	35.86%	32.73%	36.94%	36.42%	32.00%	28.78%	37.08%	34.20%	32.07%	30.25%
FNN: 20x20, 0.5x0.5, 100ep	1.86%	0.76%	7.65%	5.07%	9.33%	6.97%	4.88%	0.80%	5.80%	4.73%	6.03%	-0.78%
FNN: 40x40, 0.75x0.75, 100ep	1.00%	0.87%	5.80%	2.05%	6.16%	5.02%	0.98%	-0.62%	5.14%	2.79%	0.97%	-0.69%
FNN: 40x40, 0.5x0.5, 100ep	1.98%	0.96%	12.68%	7.51%	14.95%	9.14%	11.38%	3.59%	11.29%	7.18%	6.68%	3.19%

Model performance versus mean for different models on different data frames

- Expected linear regression to fail w/o linear correlation
- Random Forest performs the best by far
- df3 and cdf3 perform very well, df and cdf do poorly

SHAP Analysis (all performed on cdf3)

Random Forest Analysis

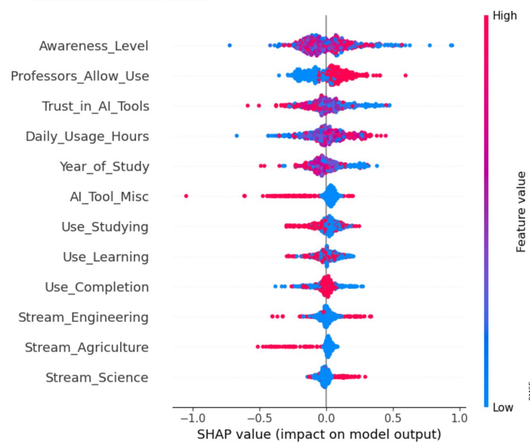


Random Forest cdf3 SHAP graph

Feature	SHAP_Importance
Daily_Usage_Hours	0.247345
Stream/Major	0.213376
Awareness_Level	0.187831
Trust_in_AI_Tools	0.157948
Professors_Allow_Use	0.145894
Year_of_Study	0.095898
Use_Cases	0.060623
Tools_Used	0.054370

Random Forest cdf3 SHAP importance

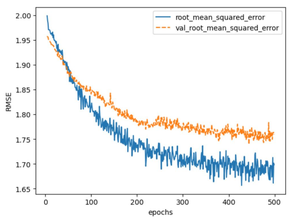
FNN Analysis (40x40 hiddens, 0.5 dropout, 500 epochs)



FNN cdf3 SHAP graph

Feature	SHAP_Importance
Stream/Major	0.213376
Awareness_Level	0.137054
Professors_Allow_Use	0.112834
Trust_in_AI_Tools	0.109431
Daily_Usage_Hours	0.107680
Year_of_Study	0.082682
AI_Tool_Misc	0.068879
Use_Studying	0.061264
Use_Cases	0.060623
Use_Learning	0.058951
Tools_Used	0.054370
Use_Completion	0.050097

FNN cdf3 SHAP importance



Summary + Conclusion

The results show that predicting how a student perceives AI's impact on their grade based on the dataset is difficult. The grade impact outliers are the hardest to predict, and limiting the dataset to students with a grade impact between -3 and 3 improves predictability greatly.

SHAP Importance tells how much a variable impacts the prediction on average, leading to some conclusions:

- ➔ Use cases tended to contribute only ~0.06 to predictions between -3 and 3; were nearly irrelevant
- ➔ Daily usage hours were important for the Random Forest Model (~0.25), but less so for the FNN (~0.10)
- ➔ Major and Awareness Level were large contributors

The SHAP Charts show whether high or low values increased the SHAP prediction or not:

- ➔ In general, more usage hours led to slightly higher perceived grade impact predictions
- ➔ Professors allowing AI usage was consistently positive
- ➔ Trust in AI tools tended to have a negative impact
- ➔ Different majors tended to have different results

Results are limited by only having the "perceived" impact, and the fact that models could only improve ~37% past guessing.

Future directions

- ➔ Machine learning model factor impact analysis seems like a good way to analyze data more complicated than linearity can account for that could continue in future research.
- ➔ A dataset that could capture a student's grade actually changing would be super useful. Additionally, I've started to suspect that a sizable portion of my dataset is generated data, which is very disappointing if true and would definitely interfere with results.
- ➔ More advanced model analysis techniques could split students into groups or evaluate how a random forest makes predictions for more interesting results

References

Abbas, Muhammad, Farooq Ahmed Jam, and Tariq Iqbal Khan. "Is It Harmful or Helpful? Examining the Causes and Consequences of Generative AI Usage Among University Students." International Journal of Educational Technology in Higher Education, vol. 21, no. 1, 2024, art. 10, https://doi.org/10.1186/s41239-024-00444-7.

Bancoro, Jaysone Christopher M. "Relationship Between Artificial Intelligence (AI) Usage and Academic Performance of Business Administration Students." Pedagogy Review: An International Journal of Educational Theories, Approaches and Strategies, vol. 1, no. 1, 2024, pp. 1-15, https://doi.org/10.62718/vmca.pr-ijetas.1.1.SC-1223-004.

Bewersdorff, Arne, et al. "AI Advocates and Cautious Critics: How AI Attitudes, AI Interest, Use of AI, and AI Literacy Build University Students' AI Self-Efficacy." Computers and Education: Artificial Intelligence, vol. 8, 2025, art. 100340, https://doi.org/10.1016/j.caeai.2024.100340.