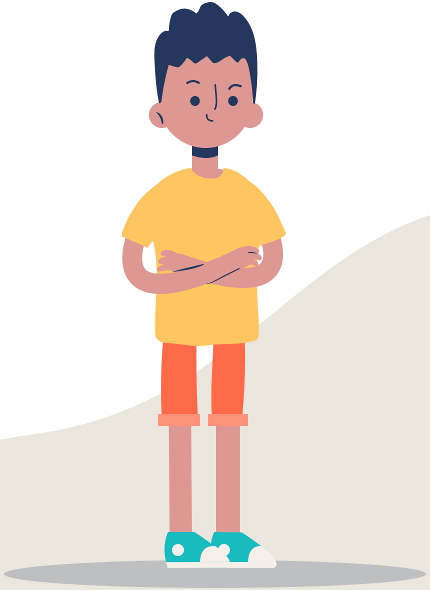
Improving Affect Detectors for K-12 Online Math Learning

Xiner Liu, Shuhan Ai, and Lisette Enumah

Penn
University of Pennsylvania

UCLA

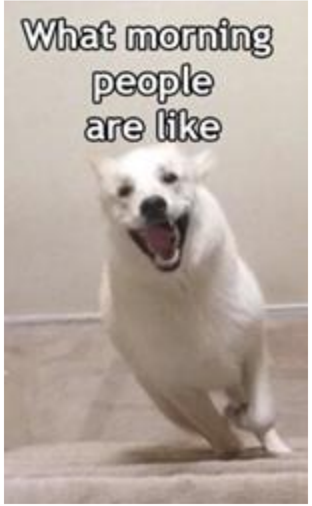


01

Research Background

Imagine your affective changes at work over one week

Early Bird vs. Night Owl



I can study all night

I want to sleep

Daytime

Night

Time Matters in Learners' Affect detection

Affect detection

Time-related variables

Previous Research

- **Affect detection** has emerged as an important area of research in the field of online education (e.g. Baker et al., 2011; Pardos et al., 2014) .
- Baker et al. (2011) developed the best **sensor-free affect detectors** for the ASSISTments platform - concentration, confusion, frustration, and boredom.

- **Time-related variables** (time of day, day of week, holidays, and season) were rarely incorporated in affect (bored, concentrating, confused, and frustrated) detection models for the ASSISTments platform.

- Only two studies touched on the concept (Pardos et al., 2014; Wang et al., 2015), and only one study included one time-related variable in the model (Wang et al. 2015).

02

Research Goals

Research Questions and Goals

Will incorporating time-related variables lead to noticeably better detection of affect?

If so, how are these variables related to affect (specifically, boredom, frustration, concentrated, and confusion) what are the implications for educators and students?

Ultimately, we want to have some actionable insights for adaptive learning platforms and instructors to better scaffold online learning.

03

Data

Source of Data

ASSISTmentsData of 2012-2013 School Data with Affect, available on ASSISTmentsData (https://sites.google.com/site/assistmentsdata/) published by Wang et. al in 2014.

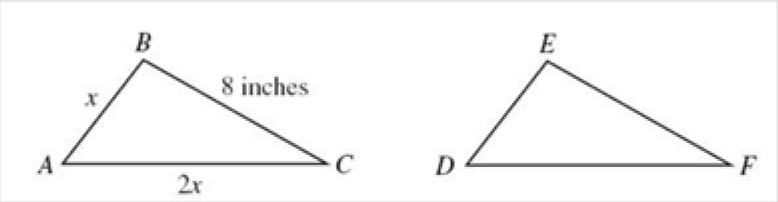
The action log files data (to be obtained) that record the start and end time of student actions within ASSISTments. The log files recorded various student actions, such as hint use, time spent, and submission information. The action log files were processed using the features engineered and detectors developed by Wang in 2015.

Source of Data - ASSISTments

You are previewing content. PRAEXE - Item 19 G-2003(Congruent triangles) (#4468)

Triangles ABC and DEF are congruent. The perimeter of triangle ABC is 23 inches.

What is the length of side DF in triangle DEF?



Break this problem into steps

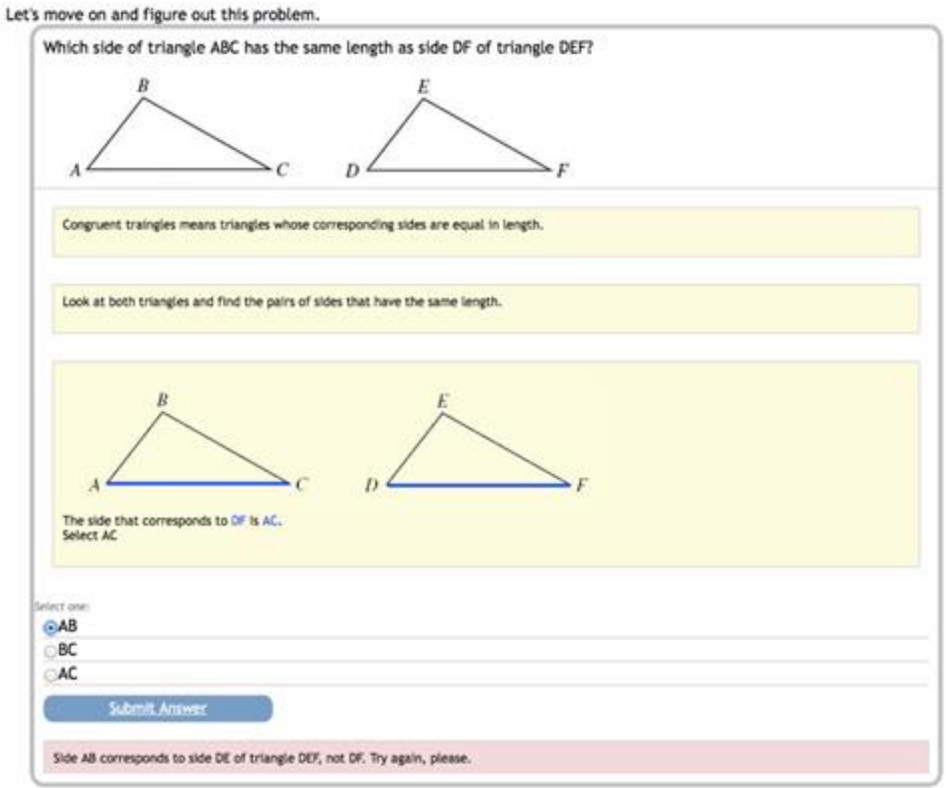
Type your answer below (mathematical expression):

5

Submit Answer

You are almost right, but remember that DF is twice x.

Source of Data - ASSISTments



Items of Interest

Features	Variables
Time of Day: Categorizes the time of day into three broad categories: morning, afternoon, and evening.	*Morning* (6:00 am - 11:59 am) *Afternoon* (12:00 pm - 5:59 pm) *Evening* (6:00 pm - 2:00 am)
Extreme Time of Day: Categorizes the extreme time periods as early morning, late night, and non-extreme hours.	*Early Morning* (4:00 am - 8:00 am) *Late Night* (9:00 pm - 4:00 am) *Non-Extreme Hours*
Day of Week: Categorize the days into weekdays and weekends.	*Weekday (Mon, Tue, Wed, Thu, Fri)* *Weekend (Sat, Sun)*
National Holidays: Categorizes the days as either 3 days before holidays, during holidays, 3 days after holidays, and other. The term 'national holidays' in this context pertains to holidays commonly celebrated in the United States, which include Independence Day, Labor Day, Thanksgiving, Christmas, and New Year's Day.	*3 Days Before Holidays* *3 Days After Holidays* *During Holidays* *Other*
Seasons: Categorizes the days into four seasons based on the calendar dates.	*Spring* (1 March - 1 June) *Summer* (1 June - 1 September) *Autumn* (1 September - 1 December) *Winter* (Dec 1 - March 1)

Most of our features will be engineered from timestamp variables in the dataset, that is, **start_time**, and **end_time** that records the **year**, **months**, and **day**, and exact time of each action.

04

Methods & Results

Exploratory Data Analysis

Data Collection

We obtain the log file from ASSISTments

Applying Detectors to Obtain Confidence Scores at Clip-Level

We applied detectors developed by **Wang et al. in 2015**, which provided confidence scores for each affective state in each clip

Average **AUC**: 0.65
Average **Kappa**: 0.24

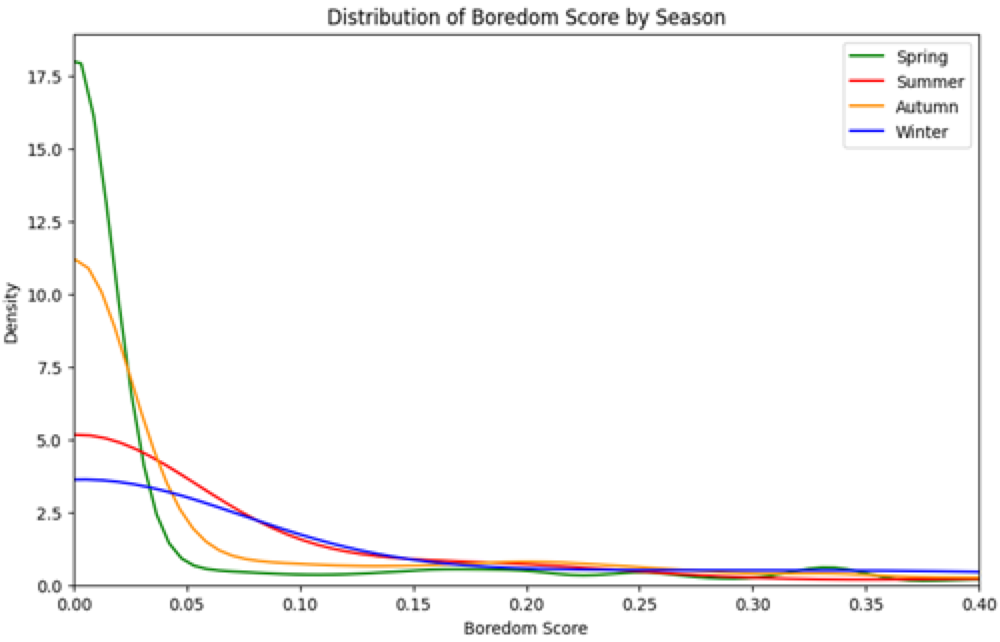
Data Visualization

We visualize the confidence scores for each affective state at various unit of time (time of the day, day of the week, season) to illustrate the impact of time-related features on student affective states.

Statistical Analysis

We conduct Kruskal-Wallis and Mann-Whitney U tests to compare affect scores across different time units to further evaluate the impact.

Exploratory Analysis: Seasonality Effect (Boredom)



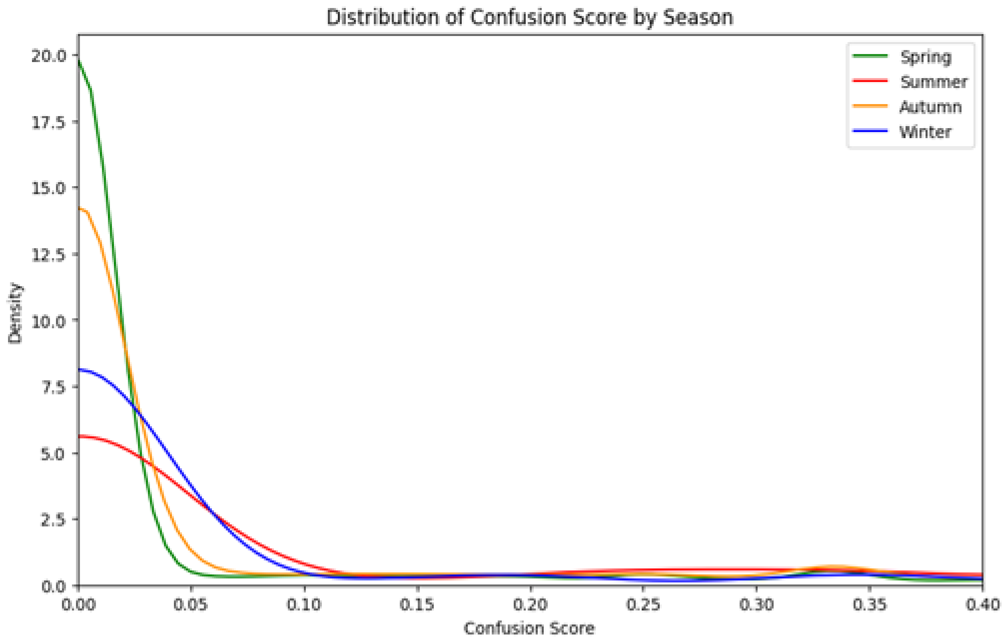
Season	25th Percentile	50th Percentile	75th Percentile	Test Statistic (H)	p-value
Spring	0.00	0.00	0.00	78.38	<0.01
Summer	0.00	0.00	0.12		
Autumn	0.00	0.00	0.10		
Winter	0.00	0.00	0.25		

Finding:

• Boredom scores were significantly **higher** during the **winter** season compared to each of the other three seasons individually ($p < 0.01$ for each test).

*** From Table 1. Statistical Results for Each Affective State by Season**

Exploratory Analysis: Seasonality Effect (Confusion)



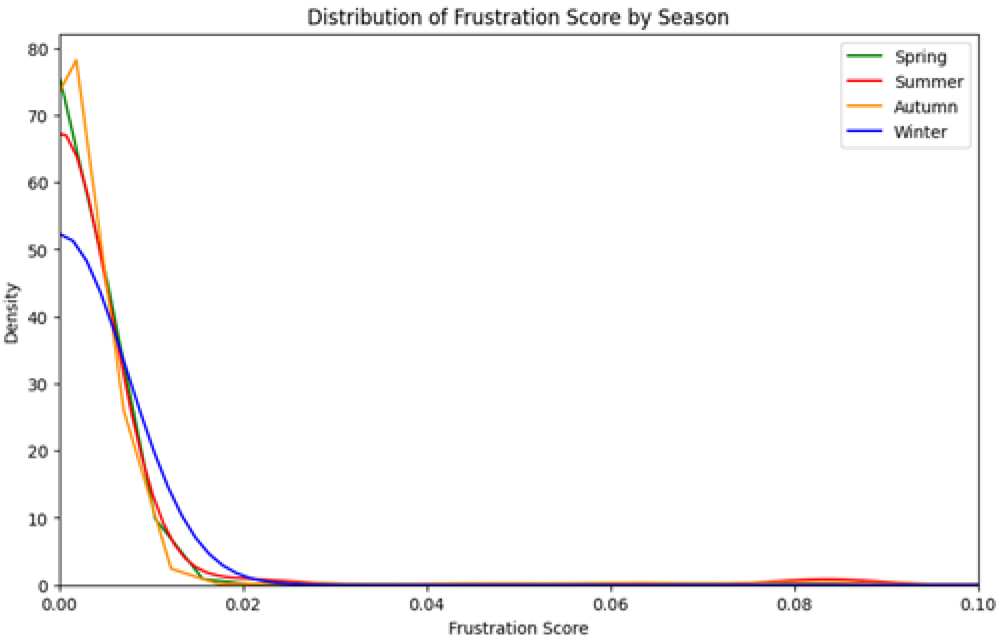
Season	25th Percentile	50th Percentile	75th Percentile	Test Statistic (H)	p-value
Spring	0.00	0.00	0.00	36.06	<0.01
Summer	0.00	0.00	0.27		
Autumn	0.00	0.00	0.00		
Winter	0.00	0.00	0.00		

Finding:

- Confusion scores were significantly **higher** during the **summer** season compared to each of the other three seasons individually ($p < 0.01$ for each test).

*** From Table 1. Statistical Results for Each Affective State by Season**

Exploratory Analysis: Seasonality Effect (Frustration)



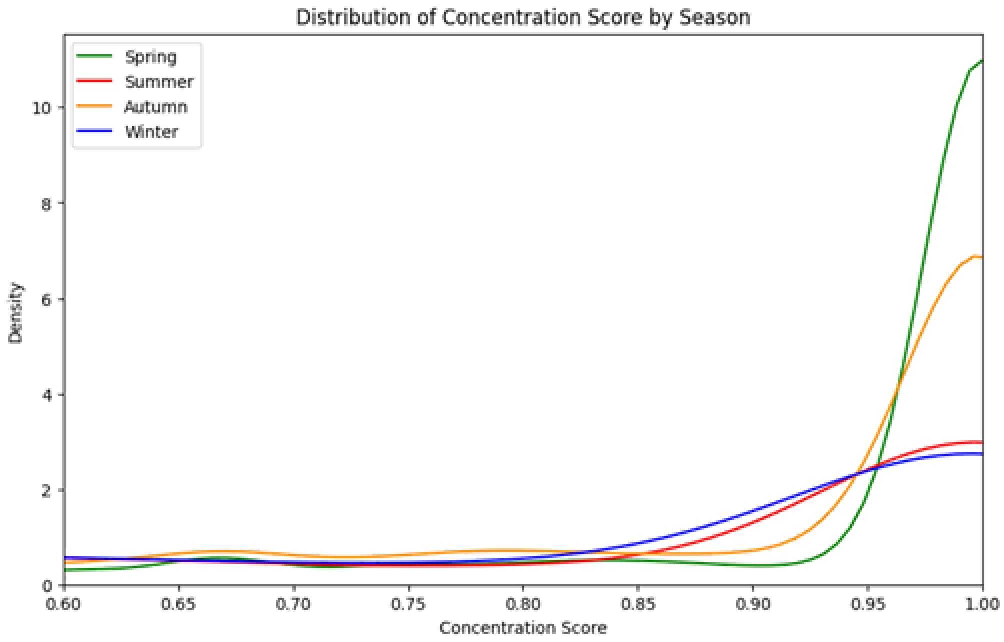
Season	25th Percentile	50th Percentile	75th Percentile	Test Statistic (H)	p-value
Spring	0.00	0.00	0.00	4.13	0.2473
Summer	0.00	0.00	0.00		
Autumn	0.00	0.00	0.00		
Winter	0.00	0.00	0.00		

Finding:

• Frustration score was found to be unaffected by seasonal changes.

*** From Table 1. Statistical Results for Each Affective State by Season**

Exploratory Analysis: Seasonality Effect (Concentration)



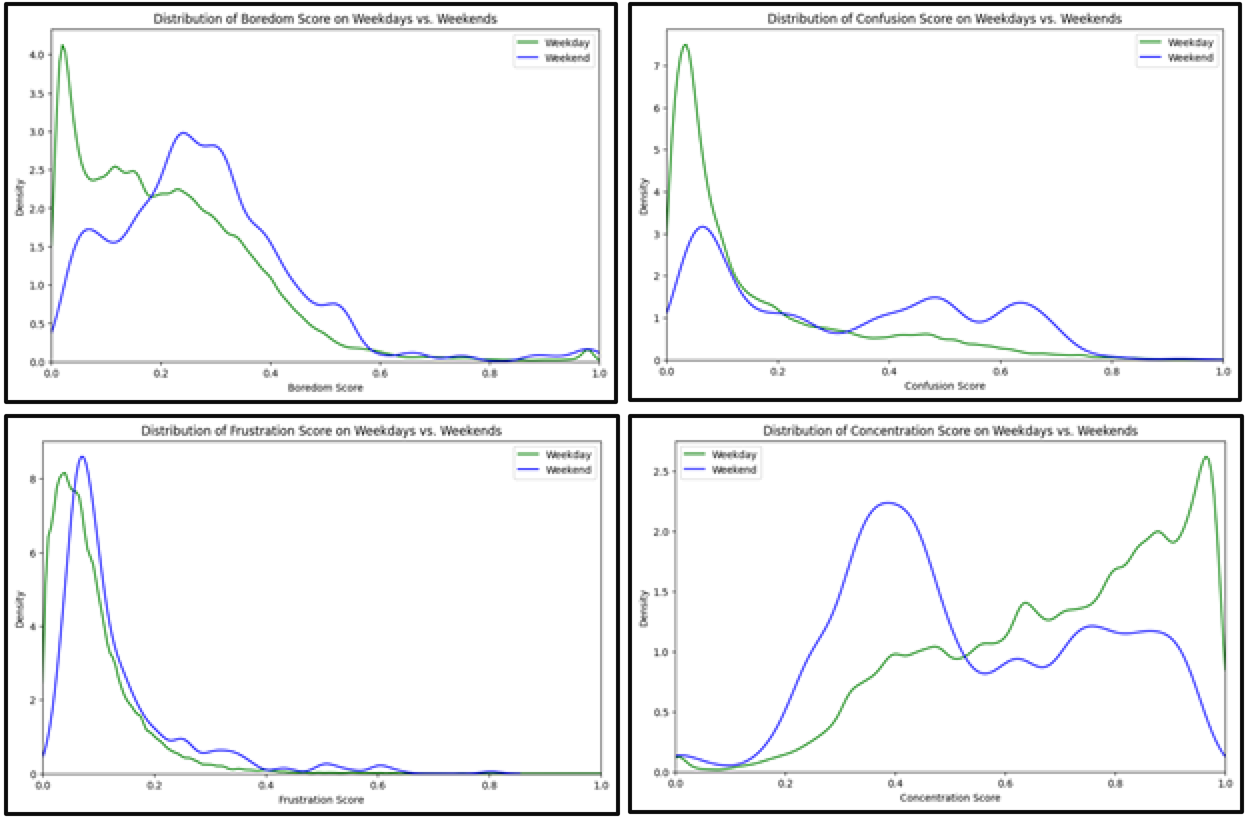
Season	25th Percentile	50th Percentile	75th Percentile	Test Statistic (H)	p-value
Spring	0.83	1.00	1.00	96.50	<0.01
Summer	0.46	1.00	1.00		
Autumn	0.67	1.00	1.00		
Winter	0.50	1.00	1.00		

Finding:

- Concentration scores were significantly **higher** during the **spring and autumn** seasons compared to the summer and winter season individually ($p < 0.01$ for each test).

*** From Table 1. Statistical Results for Each Affective State by Season**

Exploratory Analysis: Effect of Day of Week on Each Affective State



Exploratory Analysis: Effect of Day of Week on Each Affective State

Affective States	Day of Week	25th Percentile	50th Percentile	75th Percentile	Test Statistic (U)	p-value
Boredom	Weekday	0.08	0.18	0.30	5884830.5	<0.01
	Weekend	0.16	0.26	0.36		
Concentration	Weekday	0.53	0.74	0.88	11195023.5	<0.01
	Weekend	0.36	0.48	0.74		
Frustration	Weekday	0.04	0.07	0.11	5685733.0	<0.01
	Weekend	0.07	0.09	0.15		
Confusion	Weekday	0.04	0.08	0.22	4662221.5	<0.01
	Weekend	0.08	0.26	0.50		

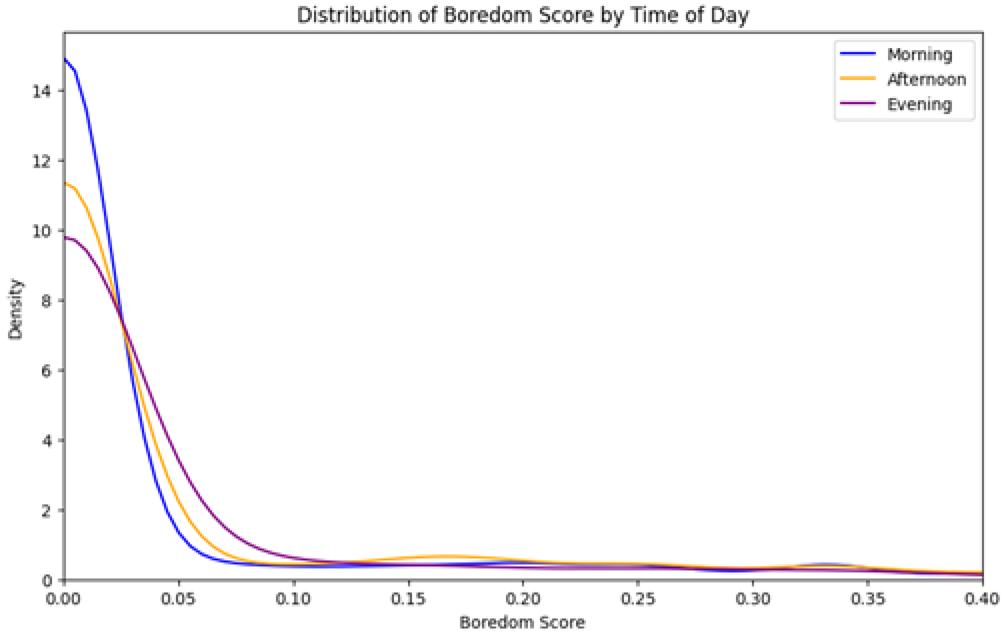
Table 2. Statistical Results for Each Affective State by Day of Week (Weekday vs. Weekend)

Finding:

• Students experienced significantly **higher** levels of **boredom, frustration, and confusion** when learning on **weekends**.

• Students experienced significantly **higher** levels of **concentration** when learning on **weekdays.**

Exploratory Analysis: Effect of Time of Day on Boredom



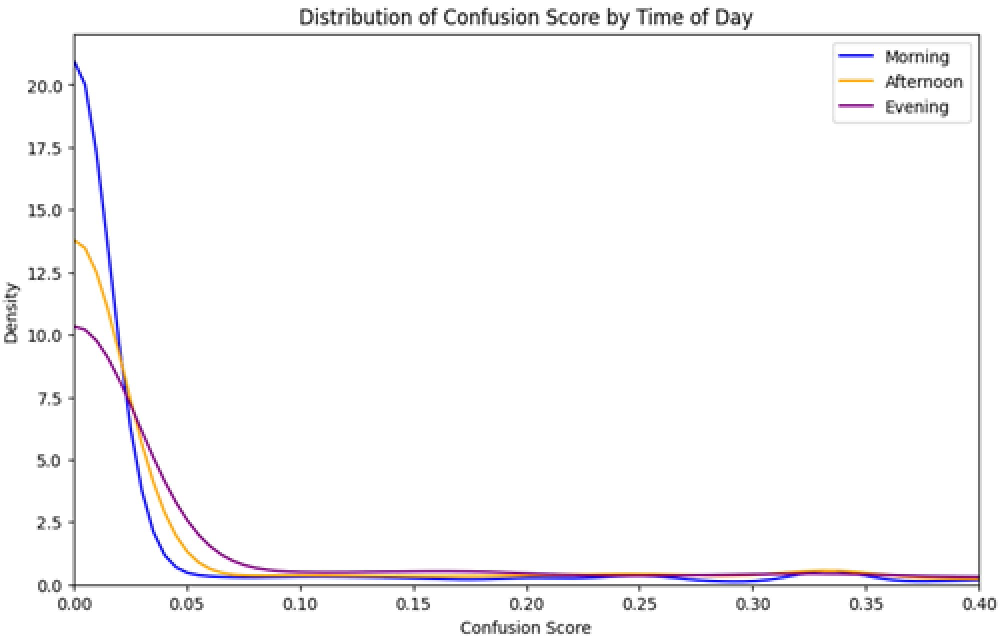
Time of Day	25th Percentile	50th Percentile	75th Percentile	Test Statistic (H)	p-value
Morning	0.00	0.00	0.00	10.56	<0.01
Afternoon	0.00	0.00	0.04		
Evening	0.00	0.00	0.00		

Finding:

- Boredom were significantly **higher** during the **evening** and **afternoon** compared to the morning individually ($p < 0.01$ for each test).

*** From Table 3. Statistical Results for Each Affective State by Time of Day.**

Exploratory Analysis: Effect of Time of Day on Confusion



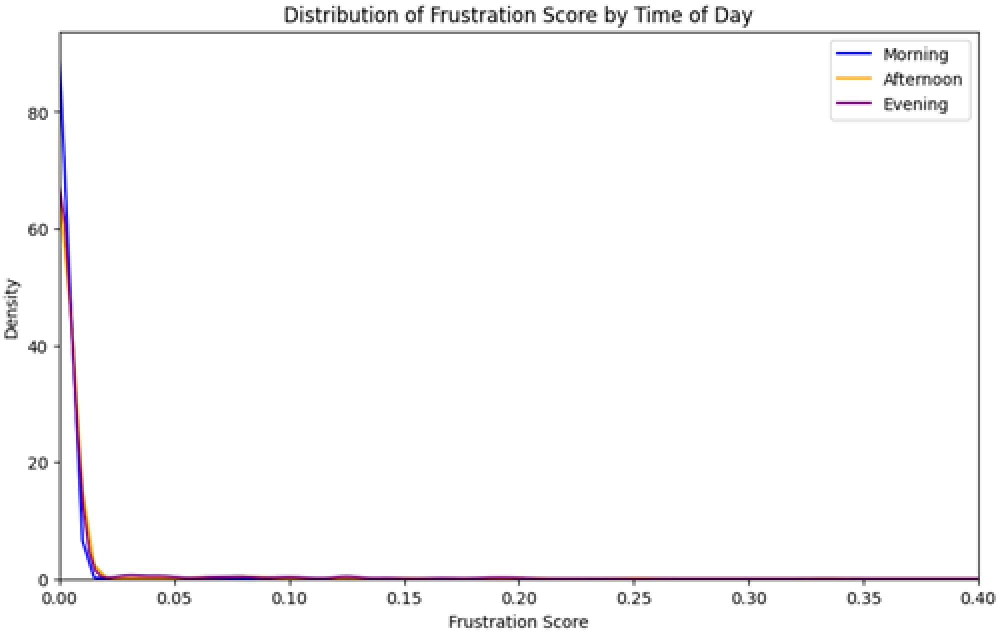
Time of Day	25th Percentile	50th Percentile	75th Percentile	Test Statistic (H)	p-value
Morning	0.00	0.00	0.00	61.52	<0.01
Afternoon	0.00	0.00	0.00		
Evening	0.00	0.00	0.06		

<u>Finding:</u>

- Confusion were significantly **higher** during the **evening** and **afternoon** compared to the morning individually (p < 0.01 for each test).

*** From Table 3. Statistical Results for Each Affective State by Time of Day.**

Exploratory Analysis: Effect of Time of Day on Frustration



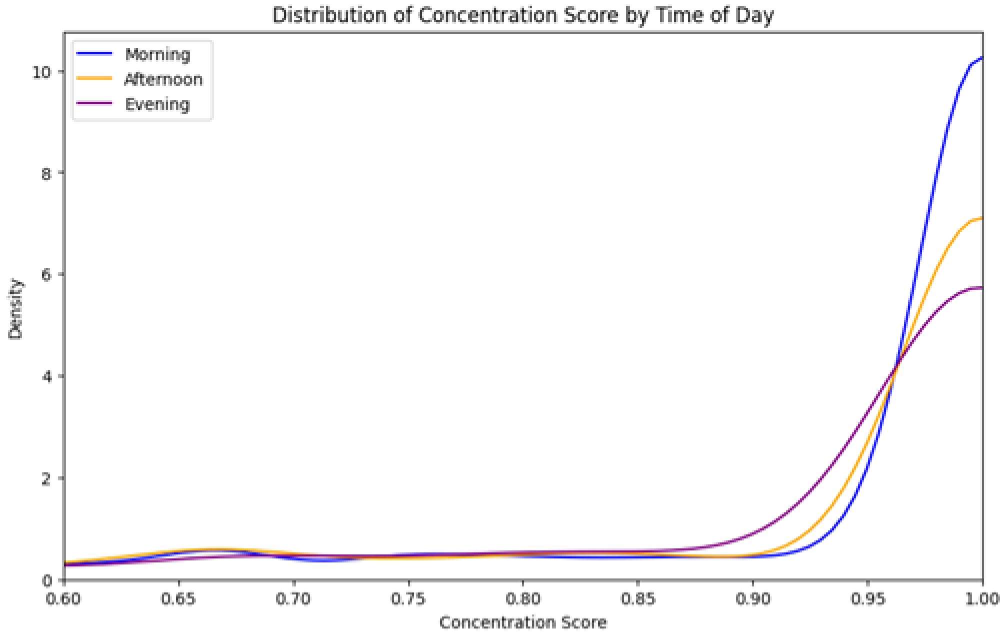
Time of Day	25th Percentile	50th Percentile	75th Percentile	Test Statistic (H)	p-value
Morning	0.00	0.00	0.00	31.19	<0.01
Afternoon	0.00	0.00	0.00		
Evening	0.00	0.00	0.00		

Finding:

• Frustration was found to be unaffected by the time of day.

*** From Table 3. Statistical Results for Each Affective State by Time of Day.**

Exploratory Analysis: Effect of Time of Day on Concentration



Time of Day	25th Percentile	50th Percentile	75th Percentile	Test Statistic (H)	p-value
Morning	0.83	1.00	1.00	36.23	<0.01
Afternoon	0.67	1.00	1.00		
Evening	0.70	1.00	1.00		

Finding:

• Concentration scores were significantly **higher** during the **morning** compared to the afternoon and evening individually (p < 0.01 for each test).

*** From Table 3. Statistical Results for Each Affective State by Time of Day.**

Time-Related Variables (Feature Engineered)

Features	Variables
Time of Day: Categorizes the time of day into three broad categories: morning, afternoon, and evening.	*Morning* (6:00 am - 11:59 am) *Afternoon* (12:00 pm - 5:59 pm) *Evening* (6:00 pm - 2:00 am)
Extreme Time of Day: Categorizes the extreme time periods as early morning, late night, and non-extreme hours.	*Early Morning* (4:00 am - 8:00 am) *Late Night* (9:00 pm - 4:00 am) *Non-Extreme Hours*
Day of Week: Categorize the days into weekdays and weekends.	*Weekday (Mon, Tue, Wed, Thu, Fri)* *Weekend (Sat, Sun)*
National Holidays: Categorizes the days as either 3 days before holidays, during holidays, 3 days after holidays, and other. The term 'national holidays' in this context pertains to holidays commonly celebrated in the United States, which include Independence Day, Labor Day, Thanksgiving, Christmas, and New Year's Day.	*3 Days Before Holidays* *3 Days After Holidays* *During Holidays* *Other*
Seasons: Categorizes the days into four seasons based on the calendar dates.	*Spring* (1 March - 1 June) *Summer* (1 June - 1 September) *Autumn* (1 September - 1 December) *Winter* (Dec 1 - March 1)

*** From Table 4. Feature Engineering for Time-related Variables.**

Feature Engineering and Model Construction

J48 Decision Tree
Step Regression
JRip
Naive Bayes
Logistic Regression
XGBoost
SVM
K*

Engineer Time-related Features

We engineered five sets of dummy-coded time-related features using the timestamp variables in the interaction log file

Construct Baseline Model*

We input features developed by Wang et al. in 2015 into eight classification algorithms to predict each affective state. Their performances served as our baseline.

Incorporate Time-related Features into the Baseline Models*

We add the time-related features into the baseline model for each affective states.

Model Evaluation

Kappa and the AUC score were used to evaluate the effectiveness of both the baseline and the new model. SHapley Additive exPlanations (SHAP) values were computed to show the relationship between the engineered features and each affective state.

*** We used 5-fold student-level cross-validation to construct each prediction models**

Time-related Variables and Model Performances

- Overall, the average **AUC** for the four affective states increased by **2%**; the average **kappa** score increased by **12.7%**

When assessing the model performance in different **sub-regions**, we found that:

- The **urban** region showed the most significant improvement in prediction accuracy (an average increase of **2.3% in AUC** & an average increase of **3.2% in kappa**)

Affective States	Region	Size	AUC ROC (Baseline Models)	AUC ROC (Models with Time-related Features)	Kappa (Baseline Models)	Kappa (Models with Time-related Features)
Engaged Concentration	All	718	0.668	**0.700**	0.335	**0.484**
	Urban	314	0.670	**0.707**	0.342	**0.490**
	Suburban	225	0.661	**0.698**	0.306	**0.474**
	Rural	179	0.656	**0.692**	0.256	**0.430**
Boredom	All	718	0.643	**0.680**	0.244	**0.257**
	Urban	314	0.648	**0.688**	0.258	**0.296**
	Suburban	225	0.646	**0.681**	0.251	**0.253**
	Rural	179	0.637	**0.672**	0.226	**0.227**
Confusion	All	718	0.651	**0.695**	0.187	**0.251**
	Urban	314	0.651	**0.701**	0.188	**0.254**
	Suburban	225	0.654	**0.698**	0.190	**0.253**
	Rural	179	0.646	**0.693**	0.180	**0.240**
Frustration	All	718	0.638	**0.660**	0.203	**0.270**
	Urban	314	0.646	**0.664**	0.210	**0.269**
	Suburban	225	0.635	**0.659**	0.198	**0.278**
	Rural	179	0.637	**0.657**	0.200	**0.263**
Average			0.666	**0.695**	0.226	**0.266**

Table 5. Effect of Time-related Features on Model Performances

The Respective Directionality of Each Feature According to the SHAP Values.

"+" represents a **positive** correlation between the feature and the target variable.

- For example: students are more likely to concentrate during the **autumn** and **spring**

"-" represents a **negative** correlation between the feature and the target variable.

- For example: they are less likely to concentrate during the **summer** and **winter**

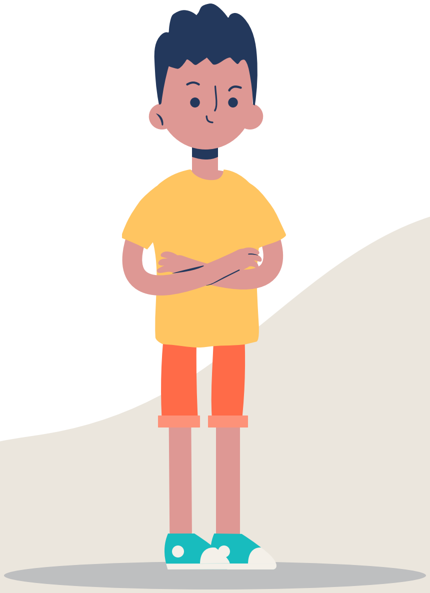
"0" represents **no correlation** between the feature and the target variable.

Engaged Concentration		Boredom		Confusion		Frustration	
Feature	SHAP Value	Feature	SHAP Value	Feature	SHAP Value	Feature	SHAP Value
Autumn	+	Evening	+	Early Morning	+	Early Morning	+
Spring	+	Afternoon	+	Non-Extreme Hours	+	Weekend	+
Early Morning	+	Winter	+	Evening	+	Afternoon	+
Weekday	+	Late Night	+	Afternoon	+	Late Night	+
Morning	+	Weekend	+	Late Night	+	3 Days After Holidays	+
Afternoon	0	Summer	+	3 Days After Holidays	+	3 Days Before Holidays	0
Not During Holidays	0	3 Days After Holidays	0	Summer	+	During Holidays	0
3 Days After Holidays	0	3 Days Before Holidays	0	Weekend	+	Not During Holidays	0
3 Days Before Holidays	0	During Holidays	0	Not During Holidays	0	Spring	0
During Holidays	0	Not During Holidays	0	3 Days Before Holidays	0	Summer	0
Weekend	-	Spring	0	Winter	0	Winter	-
Summer	-	Morning	-	During Holidays	0	Weekday	-
Winter	-	Weekday	-	Weekday	-	Evening	-
Late Night	-	Early Morning	-	Spring	-	Non-Extreme Hours	-
Non-Extreme Hours	-	Non-Extreme Hours	-	Autumn	-	Morning	-
Evening	-	Autumn	-	Morning	-	Autumn	-

Table 6. The Time-related Features in the Each Affect Detector and Their Respective Directionality According to SHAP Values.

05

Conclusion & Limitations



Discussion

Time-related features can improve model performance, particularly in the prediction of concentration and frustration.

These findings highlight the potential benefit of considering the effect of time when creating prediction models for various learning environments.

(1) Educational Games
(2) Adaptive Learning Platform
(3) Online Courses
(4) ...

Limitations and Future Directions

While this study is exploratory in nature, it has inherent limitations:

- It did not explore the use of **advanced machine learning models** such as LSTM or GNN.
- It did not study the effect of time-related features in **behavior** prediction.
- It did not test the model's **generalizability** on other users.

Future research will explore the **use of other algorithms**, the effect of time-related features in **behavior prediction**, and a **wider student population** to assess findings' generalizability

Conclusion

- Our results suggest that time-related factors (such as time of day, day of the week, and season) have an impact on student **affective states** in online learning environments.
- Incorporating these time-related features into affective state prediction models can **improve model performance**.
- This approach is also applicable to **other environments** with event-based logging systems.

Thank You