Feature Interpretability Survey

In this survey you will be asked to judge the interpretability of several different features (predictor variables) used in machine learning models to predict test-taking behavior. These features were extracted from log file records of the actions that students performed in a computer-based testing environment. The goal of this survey is to determine whether certain types of features may be more or less advantageous because of their interpretability, or lack thereof.

You will see definition of 15 features below, along with questions pertaining to each. Please make your judgments based on the definitions. After making these judgments, you will be shown the names of each feature and asked for an additional judgment based on the feature name.

# Feature Definitions

## Feature 1

Definition: Count of unusual answers (ranked outside the five most common responses) the student gave to open-ended questions in the last five minutes of the test.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 2

Definition: Count of math-related questions where the student’s time spent on the question was in the 5th percentile or higher out of all students.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 3

Definition: Sum of the squared number of seconds spent by the student per problem.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 4

Definition: Value at the 30th percentile of seconds spent by the student per problem.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 5

Definition: Standard deviation of timestamps for all events in the student’s log file from one specific problem (problem ID *VH098812*).

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 6

Definition: Slope coefficient of a linear regression that uses index in the sequence to predict time elapsed during all events in fill-in-the-blank type questions.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 7

Definition: Slope coefficient of a linear regression that uses index in the sequence to predict time elapsed for each “Math Keypress” event (working on a fill-in-the-blank type question) the student did.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 8

Definition: Mean of all sets of three consecutive values multiplied together, in the sequence of seconds spent by the student per problem.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 9

Definition: Count of “Receive focus” events (indicates starting a fill-in-the-blank type question) done by the student in the last five minutes of the test.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 10

Definition: Count of all questions (whether math-related or not) where the student’s time spent on the question was in the 5th percentile or higher out of all students.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 11

Definition: Value of the 7th coefficient from a continuous wavelet transform (the Ricker wavelet) with width 20 applied to the sequence of seconds spent by the student per problem.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 12

Definition: Skew of the distribution of the number of events in the student’s log file from one specific problem (problem ID *VH098740*).

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 13

Definition: Value at the 20th percentile of seconds spent by the student per problem.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 14

Definition: Slope coefficient of a linear regression that uses index in the sequence to predict time elapsed for all events logged for the student.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

## Feature 15

Definition: Total seconds spent on the test by the student.

How much effort does this feature definition require to understand?

[Scale: Very little (1) … A great deal (5)]

How much could you infer about a student’s learning experience from knowing that they had a particular value of this feature? (e.g., a low value, a high value)

[Scale: Very little (1) … A great deal (5)]

# Feature Names

## Feature 1

Definition: Count of unusual answers (ranked outside the five most common responses) the student gave to open-ended questions in the last five minutes of the test.

Feature name: answer\_count\_rankother\_last5

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 2

Definition: Count of math-related questions where the student’s time spent on the question was in the 5th percentile or higher out of all students.

Feature name: percentile5\_count\_actual

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 3

Definition: Sum of the squared number of seconds spent by the student per problem.

Feature name: per\_problem\_sec\_\_abs\_energy

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 4

Definition: Value at the 30th percentile of seconds spent by the student per problem.

Feature name: per\_problem\_sec\_\_quantile\_\_q\_0.3

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 5

Definition: Standard deviation of timestamps for all events in the student’s log file from one specific problem (problem ID *VH098812*).

Feature name: STD(rows.TIME\_SINCE(EventTime) WHERE AccessionNumber = VH098812)

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 6

Definition: Slope coefficient of a linear regression that uses index in the sequence to predict time elapsed during all events in fill-in-the-blank type questions.

Feature name: TREND(rows.TIME\_SINCE(EventTime), items.first\_rows\_time WHERE items.ItemType = FillInBlank)

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 7

Definition: Slope coefficient of a linear regression that uses index in the sequence to predict time elapsed for each “Math Keypress” event (working on a fill-in-the-blank type question) the student did.

Feature name: TREND(rows.TIME\_SINCE(EventTime), items.first\_rows\_time WHERE Observable = Math Keypress)

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 8

Definition: Mean of all sets of three consecutive values multiplied together, in the sequence of seconds spent by the student per problem.

Feature name: per\_problem\_sec\_\_c3\_\_lag\_1

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 9

Definition: Count of “Receive focus” events (indicates starting a fill-in-the-blank type question) done by the student in the last five minutes of the test.

Feature name: count\_Receive Focus\_last5

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 10

Definition: Count of all questions (whether math-related or not) where the student’s time spent on the question was in the 5th percentile or higher out of all students.

Feature name: percentile5\_count

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 11

Definition: Value of the 7th coefficient from a continuous wavelet transform (the Ricker wavelet) with width 20 applied to the sequence of seconds spent by the student per problem.

Feature name: per\_problem\_sec\_\_cwt\_coefficients\_\_widths\_(2, 5, 10, 20)\_\_coeff\_7\_\_w\_20

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 12

Definition: Skew of the distribution of the number of events in the student’s log file from one specific problem (problem ID *VH098740*).

Feature name: SKEW(rows.CUM\_COUNT(STUDENTID) WHERE AccessionNumber = VH098740)

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 13

Definition: Value at the 20th percentile of seconds spent by the student per problem.

Feature name: per\_problem\_sec\_\_quantile\_\_q\_0.2

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 14

Definition: Slope coefficient of a linear regression that uses index in the sequence to predict time elapsed for all events logged for the student.

Feature name: TREND(rows.TIME\_SINCE(EventTime), items.first\_rows\_time)

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

## Feature 15

Definition: Total seconds spent on the test by the student.

Feature name: sec\_spent\_total

How much of the feature definition could you have guessed based solely on its name?

[Scale: None, a little bit, approximately half, most of it, basically all of it]

# Survey Completion

You have now completed the survey. Thank you very much for lending your time and expertise!