

Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making (Supplementary Materials)

XINRU WANG and MING YIN, Purdue University

ACM Reference Format:

Xinru Wang and Ming Yin. 2021. Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making (Supplementary Materials). In *26th International Conference on Intelligent User Interfaces (IUI '21)*, April 14–17, 2021, College Station, TX, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3397481.3450650>

In this document, we provide details on the multiple-choice questions that we designed to test participants' objective understanding of the model behavior for both the recidivism prediction task (Section 1) and the forest cover prediction task (Section 2). In each question, we highlight the correct answer in red.

1 UNDERSTANDING QUESTIONS IN THE RECIDIVISM PREDICTION TASK

1.1 Feature Importance

Question 1: Among the following features, which one is the *most* important in influencing our machine learning model's prediction (that is, variations in the value of that feature will *most likely* change the model's prediction)?

- A. **Age**
- B. Charge Name
- C. Charge Degree

Question 2: Among the following features, which one is the *least* important in influencing our machine learning model's prediction (that is, variations in the value of that feature will *most unlikely* change the model's prediction)?

- A. Days in Custody
- B. Sex
- C. **Charge Degree**

1.2 Marginal effect of features on predictions

Question 3: Consider a defendant with the following profile:

1. Race	White	2. Sex	female	3. Age	31	4. Prior Count	0
5. Charge Name	Arrest case no charge			6. Detailed Charge Degree	misdemeanor	7. Days in custody	0

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2021 Copyright held by the owner/author(s).

Manuscript submitted to ACM

When all other features are kept the same,

- (1) *If the defendant's age was 41 instead of 31, how would the machine learning model's prediction on the defendant's likelihood of reoffending change?*
 - A. The model would predict the 41-year-old defendant to be *more* likely to reoffend.
 - B. **The model would predict the 41-year-old defendant to be *less* likely to reoffend.**
- (2) *If the number of days the defendant spent in custody was 100 instead of 0, how would the machine learning model's prediction on the defendant's likelihood of reoffending change?*
 - A. **The model would predict the defendant spending 100 days in the custody to be *more* likely to reoffend.**
 - B. The model would predict the defendant spending 100 days in the custody to be *less* likely to reoffend.
- (3) *If the defendant's charge name is "Driving under the Influence" instead of "Arrest with no case", how would the machine learning model's prediction on the defendant's likelihood of reoffending change?*
 - A. The model would predict the defendant who was charged with "Driving under the Influence" to be *more* likely to reoffend
 - B. **The model would predict the defendant who was charged with "Driving under the influence" to be *less* likely to reoffend**
- (4) *If the defendant's race is Hispanic instead of white, how would the machine learning model's prediction on the defendant's likelihood of reoffending change?*
 - A. The model would predict the defendant whose race is Hispanic to be *more* likely to reoffend
 - B. **The model would predict the defendant whose race is Hispanic to be *less* likely to reoffend**
- (5) *If the defendant is a male instead of a female, how would the machine learning model's prediction on the defendant's likelihood of reoffending change?*
 - A. **The model would predict the male defendant to be *more* likely to reoffend.**
 - B. The model would predict the male defendant to be *less* likely to reoffend.

1.3 Counterfactual thinking

Question 4: Consider a defendant with the following profile:

1. Race	White	2. Sex	male	3. Age	22	4. Prior Count	3
5. Charge Name	Possession of Cocaine			6. Detailed Charge Degree	felony	7. Days in custody	10

Our machine learning model currently predicts this defendant *will* reoffend. When all other features are kept the same, which of the following changes on the crime charge is **most likely** to change our model's prediction (i.e., make the model predict the defendant *will not* reoffend)?

- A. **Change the charge name to "Driving Under the Influence"**
- B. Change the charge name to "Driving with a Suspended License"
- C. Change the charge name to "Battery"

Question 5: Consider a defendant with the following profile:

1. Race	Black	2. Sex	female	3. Age	24	4. Prior Count	2
5. Charge Name	Grand Theft			6. Detailed Charge Degree	felony	7. Days in custody	117

Our machine learning model currently predicts this defendant *will* reoffend. If we change only one feature of this profile but leave all other features unchanged, which of the following changes is **going** to change our model's prediction (i.e., make the model predict the defendant *will not* reoffend)? Please check all that apply.

- A. **change Race from Black to Hispanic**
- B. change Sex from female to male
- C. **change Age from 24 to 29**
- D. **change Priors Count from 2 to 0**
- E. **change Charge Name from "Grand Theft" to "Driving Under the Influence"**
- F. **change Days in Custody from 117 to 9**

1.4 Simulate model behavior

Question 6: Consider a defendant with the following profile:

1. Race	White	2. Sex	male	3. Age	26	4. Prior Count	2
5. Charge Name	Driving Under the Influence			6. Detailed Charge Degree	felony	7. Days in custody	1

What do you think **our machine learning model will predict** for this defendant?

- A. The model will predict this defendant will reoffend within two years
- B. **The model will predict this defendant will not reoffend within two years**

Question 7: Consider three defendants with the following profiles:

Defendant 1:

1. Race	Hispanic	2. Sex	male	3. Age	41	4. Prior Count	1
5. Charge Name	Driving Under the Influence			6. Detailed Charge Degree	misdemeanor	7. Days in custody	1

Defendant 2:

1. Race	White	2. Sex	male	3. Age	24	4. Prior Count	3
5. Charge Name	Driving Under the Influence			6. Detailed Charge Degree	misdemeanor	7. Days in custody	1

Defendant 3:

1. Race	White	2. Sex	male	3. Age	25	4. Prior Count	3
5. Charge Name	Driving with a Suspended License			6. Detailed Charge Degree	misdemeanor	7. Days in custody	172

For one of these three defendants, **our machine learning model predicts** that the defendant will reoffend. Which one do you think is this defendant?

- A. Defendant 1
- B. Defendant 2
- C. **Defendant 3**

1.5 Error Detection

Question 8: Consider a defendant with the following profile:

1. Race	White	2. Sex	male	3. Age	22	4. Prior Count	0
5. Charge Name	Possession of Cocaine			6. Detailed Charge Degree	felony	7. Days in custody	1

Our machine learning model predicts that this defendant will reoffend[, and also gives its explanation on the right side]. Do you believe this prediction is correct? (w/ or w/o ML explanation)

- A. Yes, I think this prediction is correct
- B. **No, I think this prediction is wrong**

Question 9: Consider a defendant with the following profile:

1. Race	Black	2. Sex	male	3. Age	52	4. Prior Count	7
5. Charge Name	Grand Theft			6. Detailed Charge Degree	misdemeanor	7. Days in custody	1

Our machine learning model predicts that this defendant will not reoffend [, and also gives its explanation on the right side]. Do you believe this prediction is correct? (w/ or w/o ML explanation)

- A. Yes, I think this prediction is correct
 B. No, I think this prediction is wrong

2 UNDERSTANDING QUESTIONS IN THE FOREST COVER PREDICTION TASK

2.1 Feature Importance

Question 1: Among the following features, which one is the *most* important in influencing our machine learning model's prediction (that is, variations in the value of that feature will *most likely* change the model's prediction)?

- A. Elevation
 B. Slope
 C. Vertical Distance to Nearest Surface Water

Question 2: Among the following features, which one is the *least* important in influencing our machine learning model's prediction (that is, variations in the value of that feature will *most unlikely* change the model's prediction)?

- A. Aspect
 B. Hillshade Index at Noon
 C. Horizontal Distance to Nearest Roadway

2.2 Marginal effect of features on predictions

Question 3: Consider an area with the following geological profile:

1. Elevation	2500	2. Aspect	0	3. Slope	0	4. Hillshade Index at Noon	200
5. Horizontal Distance to Nearest Surface Water	0	6. Vertical Distance to Nearest Surface Water	0	7. Horizontal Distance to Nearest Roadway	0	8. Horizontal Distance to Nearest Wildfire ignition Points	0

When all other features are kept the same,

- (1) If the elevation was 3000m instead of 2500m, how would the machine learning model's prediction on the likelihood of primary tree species being spruce/fir change?
 A. The model would predict the primary tree species is *more* likely to be spruce/fir.
 B. The model would predict the primary tree species is *less* likely to be spruce/fir.
- (2) If the aspect was 100° instead of 0°, how would the machine learning model's prediction on the likelihood of primary tree species being spruce/fir change?
 A. The model would predict the primary tree species is *more* likely to be spruce/fir.
 B. The model would predict the primary tree species is *less* likely to be spruce/fir.
- (3) If the slope was 10° instead of 0°, how would the machine learning model's prediction on the likelihood of primary tree species being spruce/fir change?
 A. The model would predict the primary tree species is *more* likely to be spruce/fir.

- B. The model would predict the primary tree species is *less likely to be spruce/fir*.
- (4) If the hillshade index at noon was 250 instead of 200, how would the machine learning model's prediction on the likelihood of primary tree species being spruce/fir change?
- A. The model would predict the primary tree species is *more likely to be spruce/fir*.
- B. The model would predict the primary tree species is *less likely to be spruce/fir*.
- (5) If the Horizontal Distance to Nearest Surface Water was 1000m instead of 0m, how would the machine learning model's prediction on the likelihood of primary tree species being spruce/fir change?
- A. The model would predict the primary tree species is *more likely to be spruce/fir*.
- B. The model would predict the primary tree species is *less likely to be spruce/fir*.

2.3 Counterfactual thinking

Question 4: Consider an area with the following geological profile:

1. Elevation	3247	2. Aspect	51	3. Slope	25	4. Hillshade Index at Noon	176
5. Horizontal Distance to Nearest Surface Water	242	6. Vertical Distance to Nearest Surface Water	77	7. Horizontal Distance to Nearest Roadway	4073	8. Horizontal Distance to Nearest Wildfire ignition Points	3220

Our machine learning model currently predicts the primary tree species in this area *is* spruce/fir. When all other features are kept the same, which of the following changes on the Elevation is **most likely** to change our model's prediction (i.e., make the model predict the primary tree species in this area *is not* spruce/fir)?

- A. Change the Elevation to 2799
- B. Change the Elevation to 3801

Question 5: Consider an area with the following geological profile:

1. Elevation	3224	2. Aspect	314	3. Slope	15	4. Hillshade Index at Noon	227
5. Horizontal Distance to Nearest Surface Water	42	6. Vertical Distance to Nearest Surface Water	9	7. Horizontal Distance to Nearest Roadway	3293	8. Horizontal Distance to Nearest Wildfire ignition Points	4142

Our machine learning model currently predicts the primary tree species in this area *is* spruce/fir. If we change only one feature of this profile but leave all other features unchanged, which of the following changes is **going** to change our

model's prediction (i.e., make the model predict the primary tree species in this area *is not* spruce/fir)? Please check all that apply.

- A. **change Elevation from 3224 to 2982**
- B. **change Aspect from 314 to 11**
- C. **change Slope from 15 to 28**
- D. **change Hillshade Index at Noon from 227 to 247**
- E. **change Horizontal Distance to Nearest Surface Water from 42 to 632**
- F. change Vertical Distance to Nearest Surface Water from 9 to 104
- G. change Horizontal Distance to Nearest Roadway from 3293 to 2012
- H. change Horizontal Distance to Nearest Wildfire ignition Points from 4142 to 213

2.4 Simulate model behavior

Question 6: Consider an area with the following geological profile:

1. Elevation	2648	2. Aspect	188	3. Slope	28	4. Hillshade Index at Noon	247
5. Horizontal Distance to Nearest Surface Water	376	6. Vertical Distance to Nearest Surface Water	218	7. Horizontal Distance to Nearest Roadway	362	8. Horizontal Distance to Nearest Wildfire ignition Points	860

What do you think **our machine learning model will predict** for this area?

- A. The model will predict primary tree species in this area is spruce/fir
- B. **The model will predict primary tree species in this area is not spruce/fir**

Question 7: Consider three areas with the following geological profiles:

Area 1:

1. Elevation	2341	2. Aspect	26	3. Slope	25	4. Hillshade Index at Noon	177
5. Horizontal Distance to Nearest Surface Water	242	6. Vertical Distance to Nearest Surface Water	75	7. Horizontal Distance to Nearest Roadway	330	8. Horizontal Distance to Nearest Wildfire ignition Points	1006

Area 2:

1. Elevation	3290	2. Aspect	217	3. Slope	16	4. Hillshade Index at Noon	254
5. Horizontal Distance to Nearest Surface Water	511	6. Vertical Distance to Nearest Surface Water	154	7. Horizontal Distance to Nearest Roadway	912	8. Horizontal Distance to Nearest Wildfire ignition Points	4589

Area 3:

1. Elevation	3247	2. Aspect	51	3. Slope	25	4. Hillshade Index at Noon	176
5. Horizontal Distance to Nearest Surface Water	242	6. Vertical Distance to Nearest Surface Water	77	7. Horizontal Distance to Nearest Roadway	4073	8. Horizontal Distance to Nearest Wildfire ignition Points	3220

For one of these three areas, **our machine learning model predicts** that the primary tree species in this area is spruce/fir. Which one do you think is this area?

- A. Area 1
- B. Area 2
- C. **Area 3**

2.5 Error Detection

Question 8: Consider an area with the following geological profile:

1. Elevation	3160	2. Aspect	0	3. Slope	14	4. Hillshade Index at Noon	213
5. Horizontal Distance to Nearest Surface Water	771	6. Vertical Distance to Nearest Surface Water	143	7. Horizontal Distance to Nearest Roadway	1766	8. Horizontal Distance to Nearest Wildfire ignition Points	2451

Our machine learning model predicts that the primary tree species in this area is not spruce/fir[, and also gives its explanation on the right side]. Do you believe this prediction is correct? (w/ or w/o explanation)

A. Yes, I think this prediction is correct

B. **No, I think this prediction is wrong**

Question 9: Consider an area with the following geological profile:

1. Elevation	2868	2. Aspect	347	3. Slope	4	4. Hillshade Index at Noon	232
5. Horizontal Distance to Nearest Surface Water	30	6. Vertical Distance to Nearest Surface Water	0	7. Horizontal Distance to Nearest Roadway	2947	8. Horizontal Distance to Nearest Wildfire ignition Points	2137

Our machine learning model predicts that the primary tree species in this area is spruce/fir[, and also gives its explanation on the right side]. Do you believe this prediction is correct? (w/ or w/o explanation)

A. Yes, I think this prediction is correct

B. **No, I think this prediction is wrong**