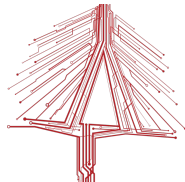
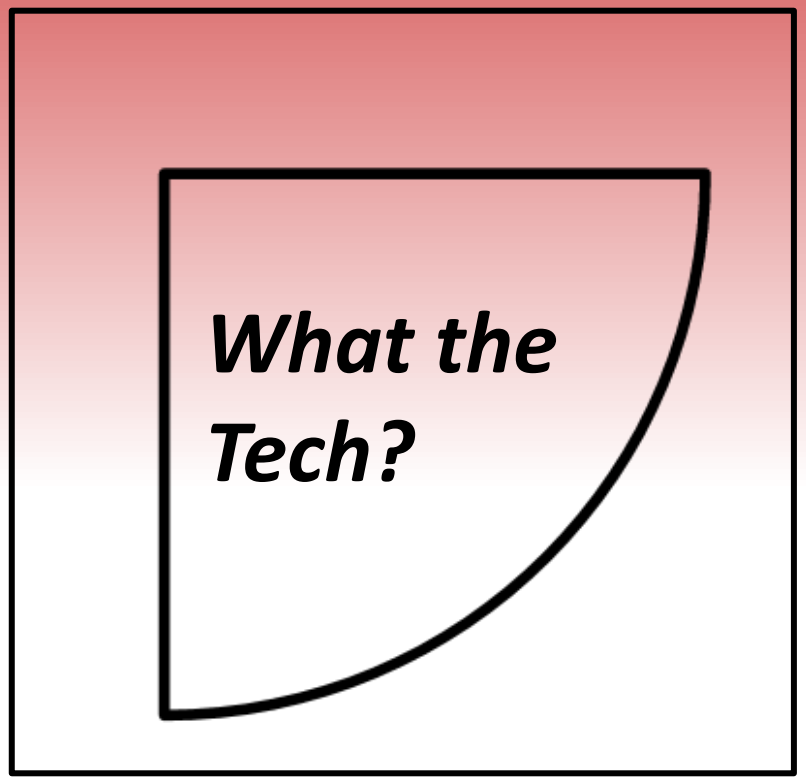
What the Tech?

PART 4:

PREDICTIVE ANALYTICS AND AI



PREDICTIVE ANALYTICS AND AI

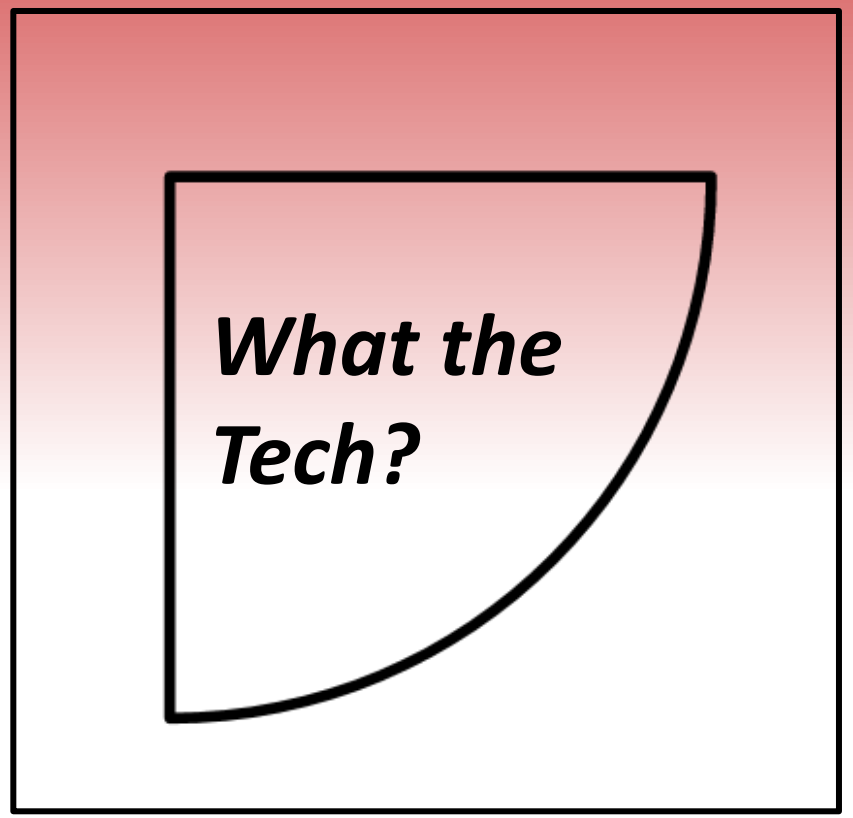


Last time, we…

- Shared the second round of our reflective assignments
- Talked about Ethics
  - Overview
  - In AI (different settings)
  - Small group activity

Any questions? Refreshers?

PREDICTIVE ANALYTICS AND AI



Today will be about:

- Third round of reflective assignments
- Predictive Analytics & AI
  - Definition
  - The data it uses
  - Where we see it today
- Demonstration activity using KNIME

Questions so far?

Next: Our Weekly Reflections

Weekly Reflections

How it works

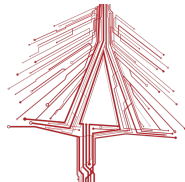
- Goal: everyone goes once!
- Three volunteers per week
- Three new volunteers each week

Every Monday

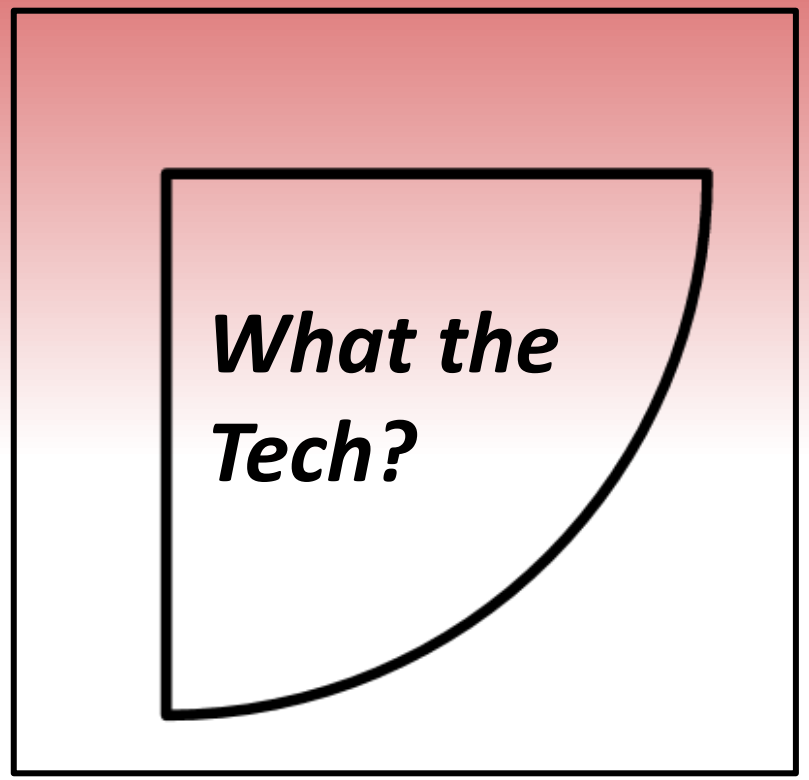
- Week 2:
- Week 3:
- Week 4: **Today**
- Week 5:
- Week 7:

Questions so far?

Next: PREDICTIVE ANALYTICS AND AI



PREDICTIVE ANALYTICS AND AI



So far, we discussed…

- What AI is
  - Definition and history
- The components and keywords of AI
- The ethics surrounding AI's application and usage

What The Tech? - The Story So Far

Part One: The Basics

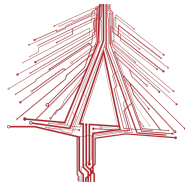
- What AI is
  - Its definition and history
- The components and keywords of AI
  - Input, output, and everything in between
- The ethics surrounding AI's application and usage
  - Deep fakes, fake news, etc.

Part Two: Different AI Types

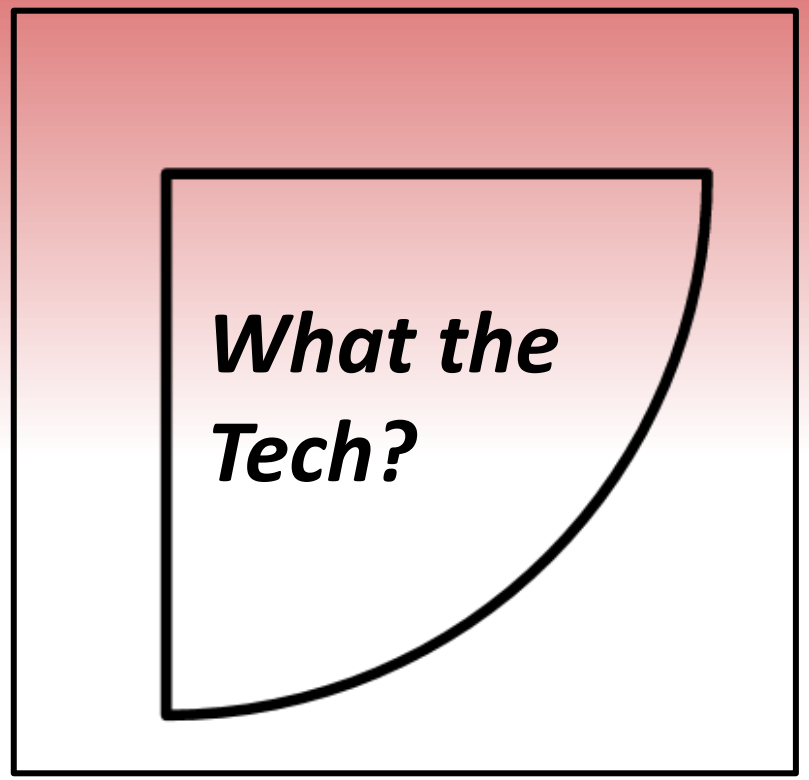
- **Predictive Analytics and AI (today)**
- Generative AI
- AI in Robotics and Automation

Part Three: AI in the Everyday

- AI use and governance in our communities
- Final project prep and presentation



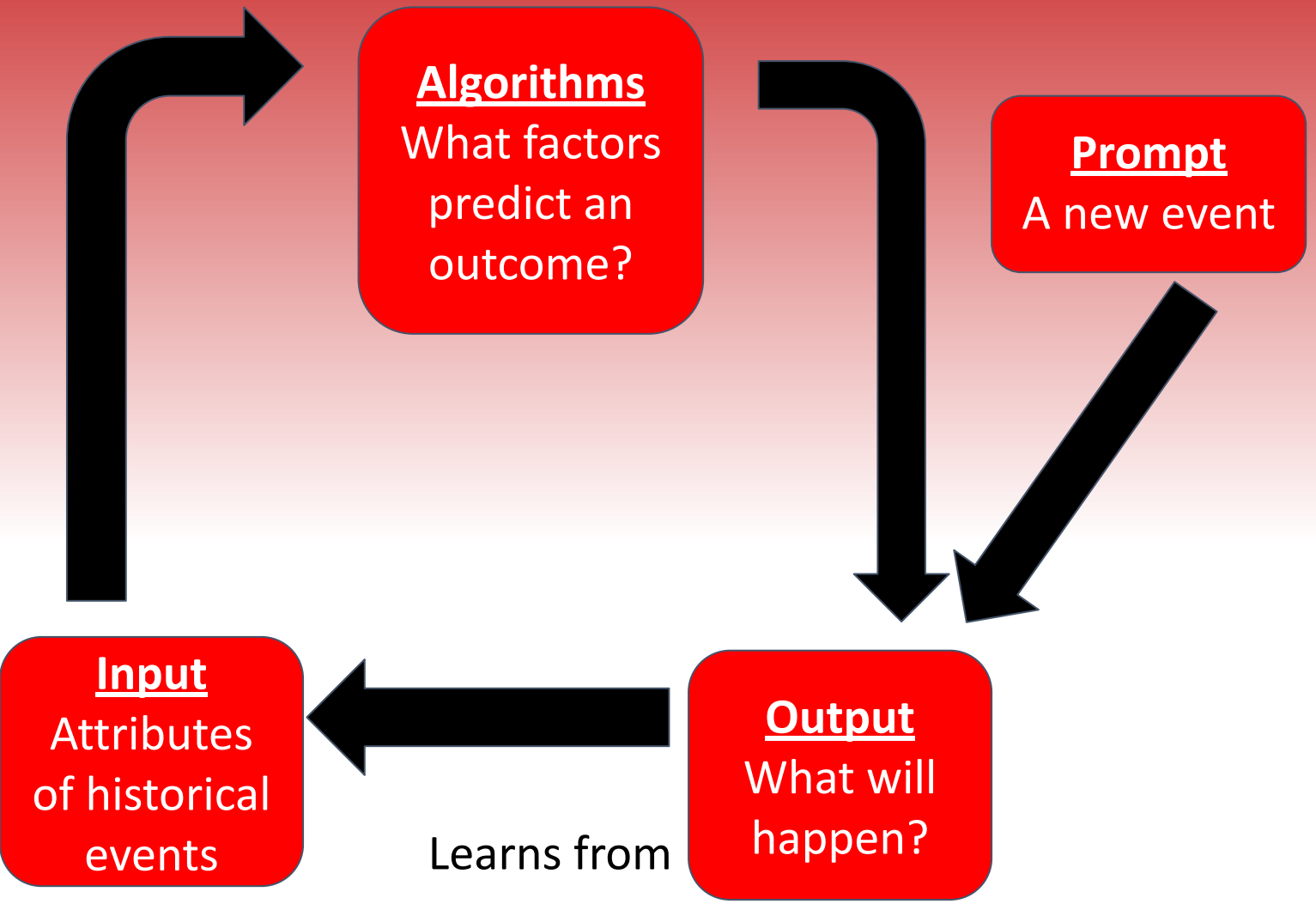
PREDICTIVE ANALYTICS AND AI



What *are* predictive analytics?

- Predictive analytics is the use of historical data to predict future events or situations
- This is done through “machine learning”
  - Discussed this in Week 2 (“Layers of AI”)
  - Who remembers the definition?

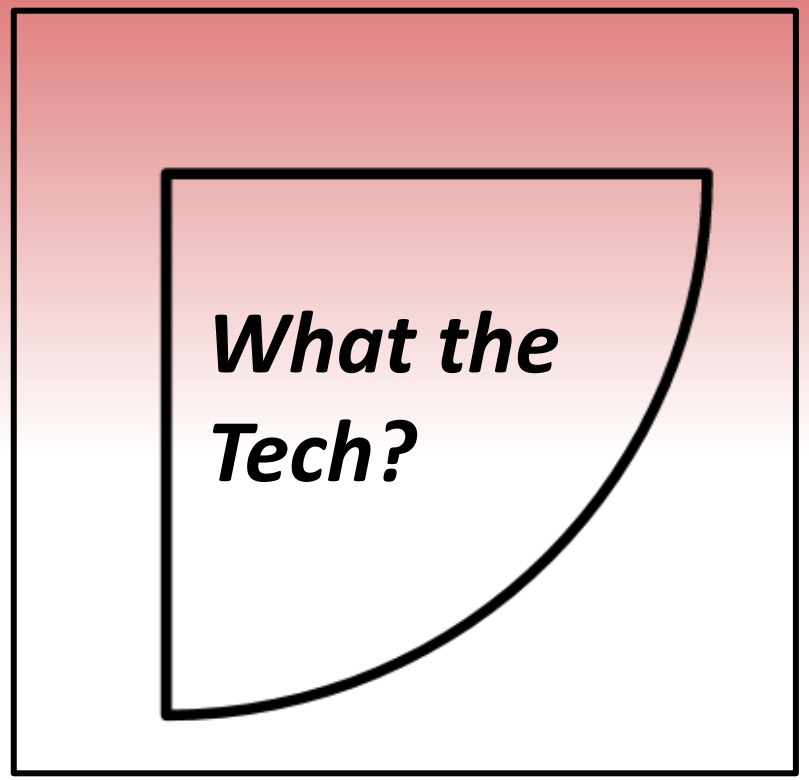
PREDICTIVE ANALYTICS AND AI



Machine learning is…

- When an algorithm generates a model using training data with examples of correct decisions.
- Algorithm uses the model it produced in training to make predictions from new data.
  - This is what we mean when we say “the algorithm recommended X”

PREDICTIVE ANALYTICS AND AI

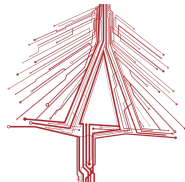


Examples of Historical Data:

- HR using hiring and firing data to predict future employee performance
- Spotify, Netflix, etc. use your streaming history to inform business decisions
- Occurrences and frequencies of certain crimes
- Sports teams predicting the attributes of a successful player.
  - This will be our focus today

Questions so far?

Next: Demonstrating Predictive Analytics & AI with KNIME



NBA Basketball Analysis

- Most used tools for Predictive Analytics
- Quick Intro to KNIME
- Example: Collecting and analyzing 10 years of NBA Data
- Can we predict player performance?

Tools for Predictive Analytics

Open Source

- Python : several packages including TensorFlow (Google) , PyTorch (Facebook)
- R
- KNIME

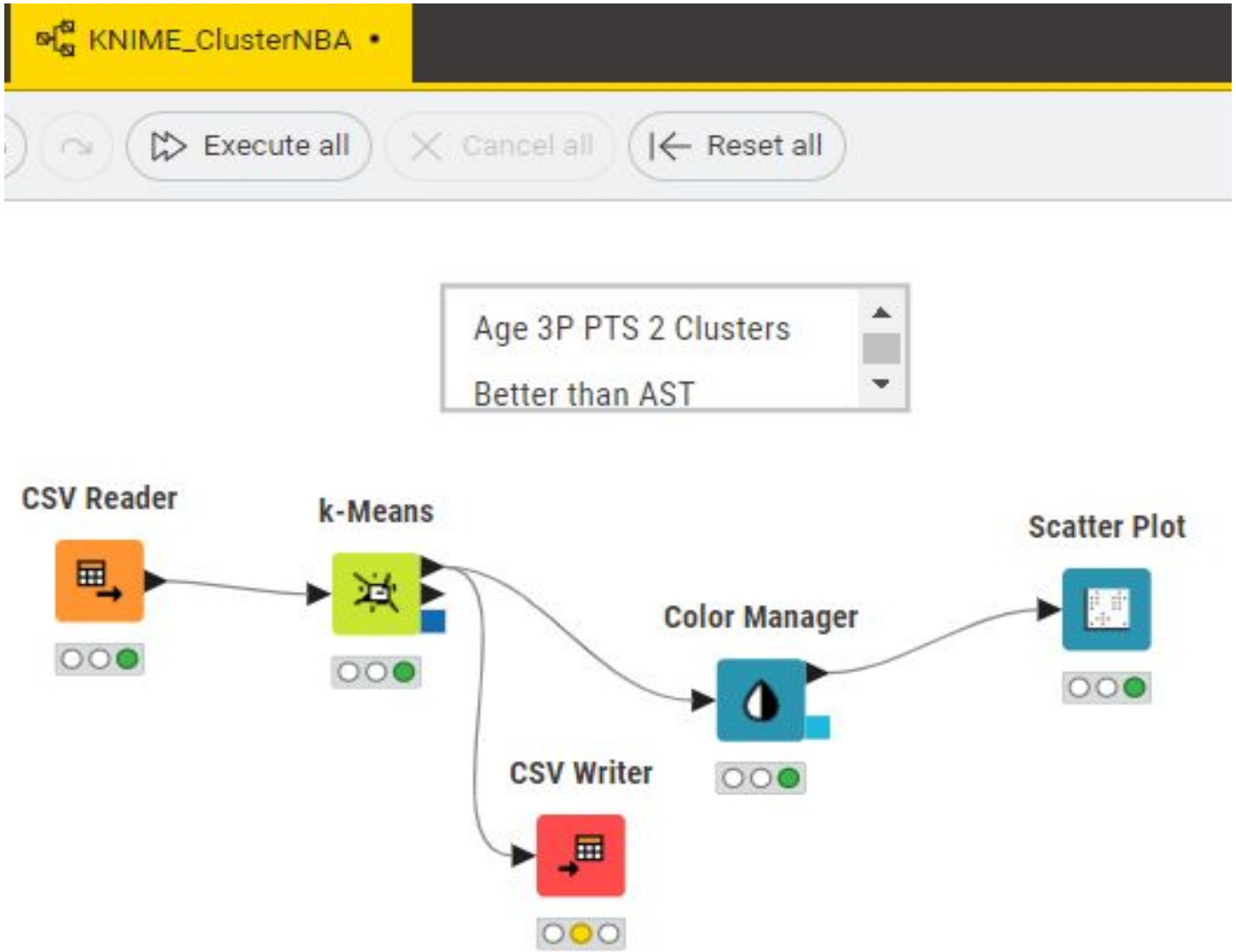
Proprietary

- IBM Watson Studio
- Microsoft Azure Machine Learning
- Google Cloud AI Platform
- Amazon SageMaker
- DataRobot
- H2O.ai (also an Open-Source version)
- Databricks

What is KNIME?

- KNIME is a drag and drop tool used for statistical analysis and machine learning.
- Imagine you have a bunch of ingredients (data) that you want to cook up into a delicious dish (insights).
- KNIME helps you do that by providing all sorts of tools and gadgets (nodes) that you can use to prepare, mix, and analyze your ingredients.
- You start with raw data, just like you start with raw ingredients for cooking. Then, you use KNIME to chop, dice, mix, and transform your data just like a chef would prep ingredients.
- Once everything is ready, you can use KNIME to bake, fry, or boil your data (analyze, visualize, and model) until you get the perfect result (insights).

KNIME for AI

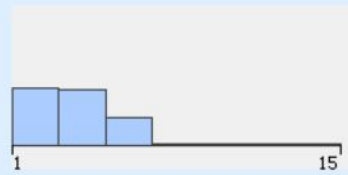
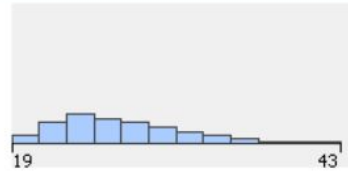
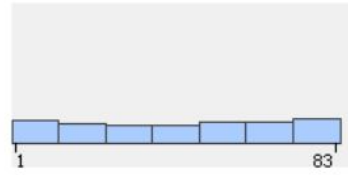
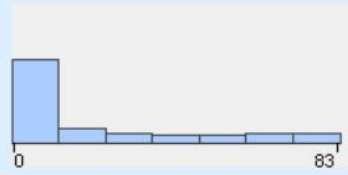
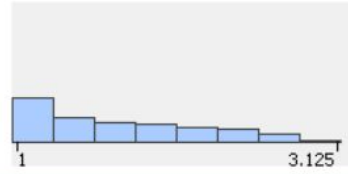
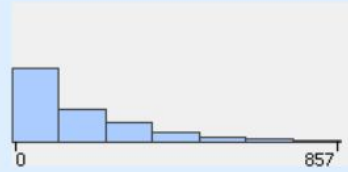


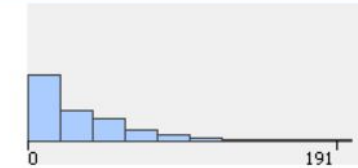
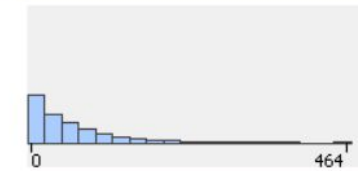
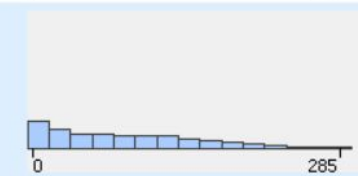
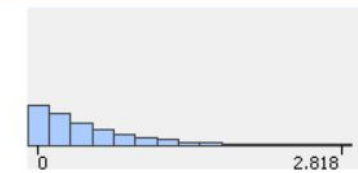
Open for Innovation

KNIME

Glossary of attributes

Descriptive Statistics

Column	Min	Mean	Median	Max	Std. Dev.	Skewness	Kurtosis	No. Missing	No. +∞	No. -∞	Histogram
Pos	1	3.0708	?	15	1.6624	1.6314	8.2073	0	0	0	
Age	19	26.3133	?	43	4.2287	0.6186	-0.1187	0	0	0	
G	1	42.2692	?	83	25.2641	-0.0758	-1.2944	0	0	0	
GS	0.0	19.4549	?	83	25.2894	1.1574	-0.0784	0	0	0	
MP	1	952.2241	?	3,125	788.4244	0.578	-0.791	0	0	0	
FG	0.0	155.2223	?	857	156.0172	1.2874	1.3268	0	0	0	

Column	Min	Mean	Median	Max	Std. Dev.	Skewness	Kurtosis	No. Missing	No. +∞	No. -∞	Histogram
AST	0.0	91.6696	?	907	112.1055	2.1597	5.7874	0	0	0	
STL	0.0	29.9639	?	191	29.2043	1.3685	2.0842	0	0	0	
BLK	0.0	18.9296	?	219	25.7304	2.8195	10.766	0	0	0	
TOV	0.0	52.5897	?	464	54.0444	1.6028	3.22	0	0	0	
PF	0.0	79.0879	?	285	63.0752	0.6003	-0.5132	0	0	0	
PTS	0.0	420.7941	?	2,818	430.4609	1.4044	1.9329	0	0	0	
Year	1,314	1,822.4651	?	2,324	321.0398	-0.0057	-1.2261	0	0	0	

KMEANS in KNIME

Imagine you have a big basket of colored marbles, but they're all mixed up. Your job is to group them based on their color. The KMeans algorithm is like a smart sorting machine that helps you do this automatically.

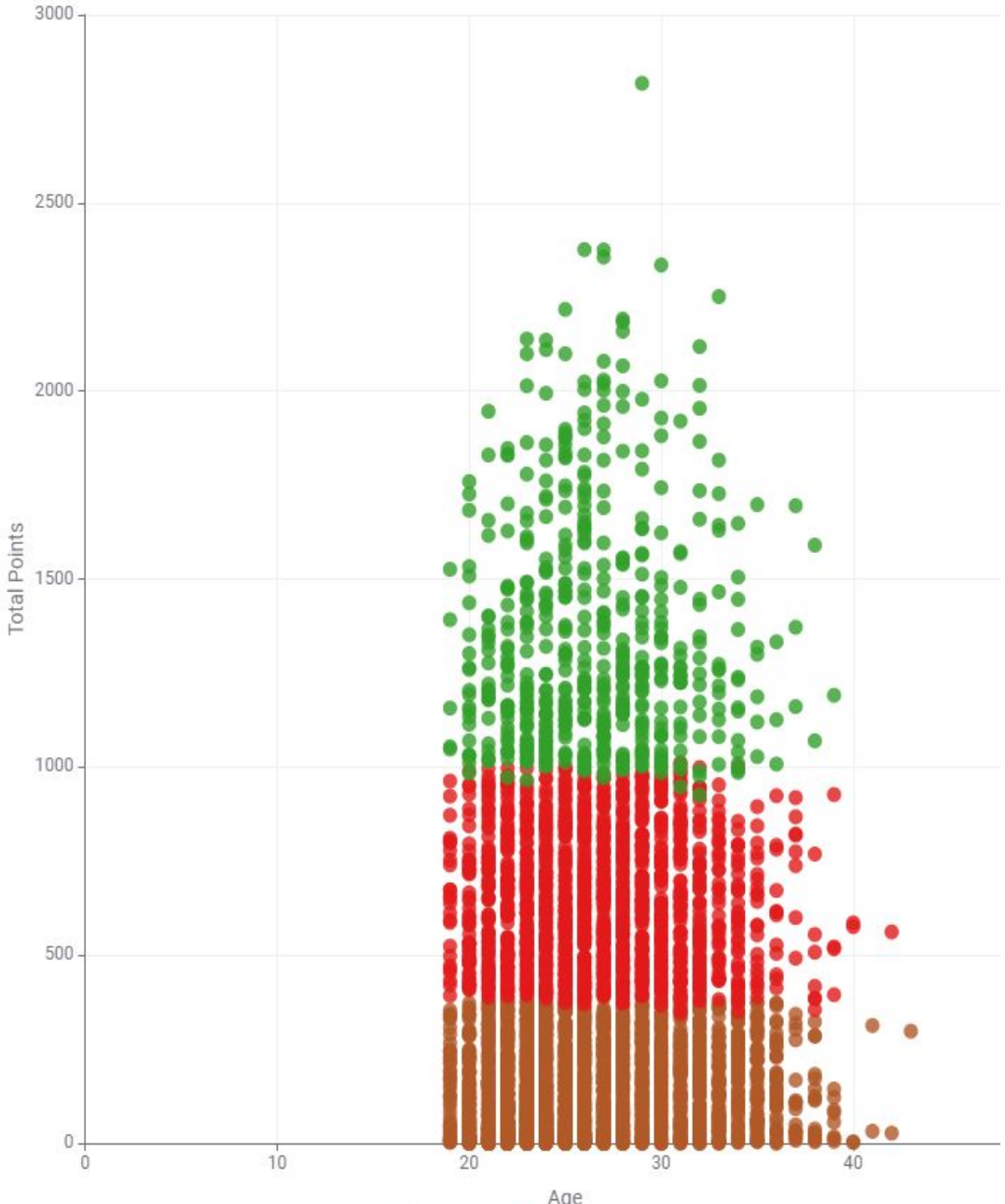
1. **Put the Marbles in the Machine:** You load all the marbles (data points) into the KNIME KMeans machine.
2. **Decide on Groups:** You tell the machine how many groups (clusters) you want, like saying you want to sort the marbles into three color groups.
3. **Sorting Begins:** The machine starts sorting the marbles based on their color similarity. It tries to make groups where marbles within each group are as similar as possible in color.
4. **Repeat Until Sorted:** The machine keeps rearranging the marbles until it finds the best way to group them based on the colors.

After the machine finishes, you'll have your marbles neatly sorted into color groups! KMeans does something similar but with data points instead of marbles, making it easier to understand patterns and relationships in the data.

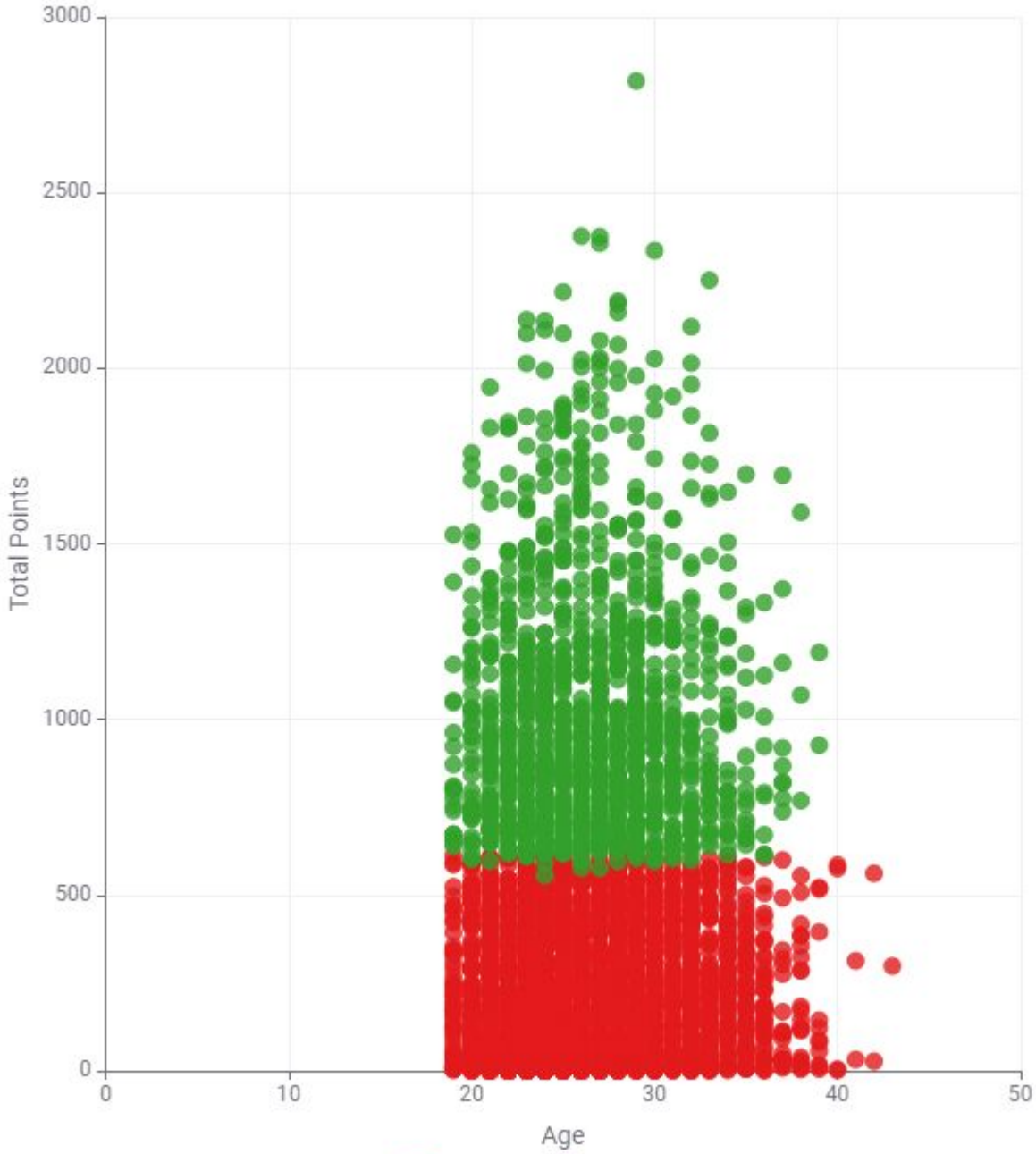
NBA Clustering in KNIME

- Can we predict how good a player is?
- We don’t have any label on our data
- We could try to sort all the players into two or three categories
- Which categories could we think off?
- How do we do that using KNIME?

Brown: Good Red:Excellent Green: Outstanding



Scatter Plot Red: Good Green:Excellent



Creating Rules based on the clusters

- Could a Tree Model help us predict if a player is outstanding?
- Why is the prediction not so straightforward?
- How much data is enough?
- What should we consider when using historical data?

What is a classification tree model?

You're planning a vacation but you're not sure where to go. You have a list of destinations and criteria that are important to you, like weather, activities, and cost. The KNIME tree model is like a super-smart travel advisor that helps you make the best decision based on your criteria.

1. **Destination Options:** You start with a list of vacation destinations, each represented by a different branch of the tree.
2. **Decision Criteria:** You have criteria like "Sunny weather," "Outdoor activities," and "Affordable prices." These criteria become the decision points in the tree.
3. **Branching Decisions:** As you make decisions (answer criteria questions), the tree branches out to different destinations based on your choices. For example, if you choose "Sunny weather," the tree might branch to locations known for sunny days.
4. **Final Recommendation:** After answering all the criteria questions, the tree leads you to a final recommendation—a vacation destination that meets all your criteria.

Imagine a tree upside down; the travel is the trunk; destination options are branches and each final decision is a leaf.

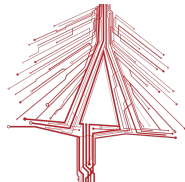
Your time to play with those tools!

It is a pleasure to get you started with Predictive Analytics and AI.

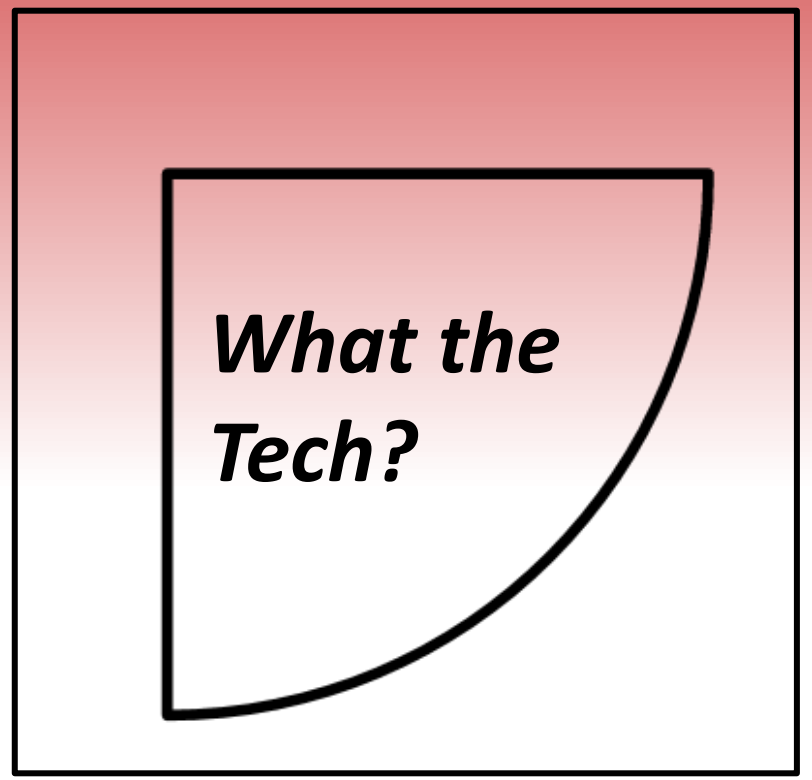
Thank you !!!

Alice
Mello

Questions?



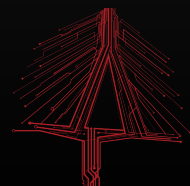
PREDICTIVE ANALYTICS AND AI



Next time:

- Continue Predictive AI discussion
  - Terms
  - Group activity: Teachable Machine (Google)
  - Ethics
- Talk about the final project, etc.
- Debrief
- Look ahead at Week 5
  - Generative AI

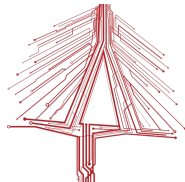
HALFWAY POINT



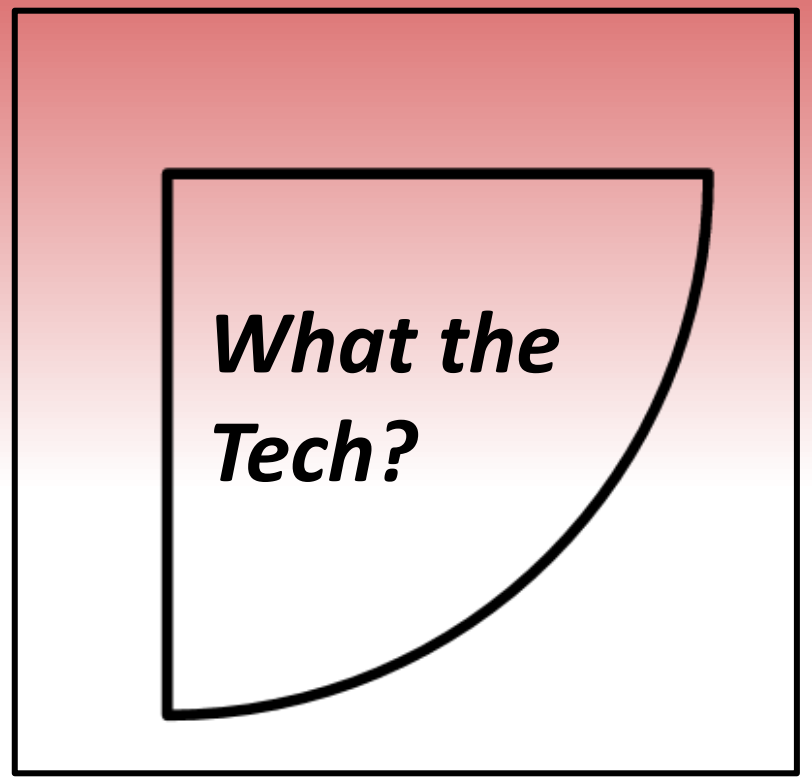
What the Tech?

Week 4, Part 2:

PREDICTIVE ANALYTICS AND AI



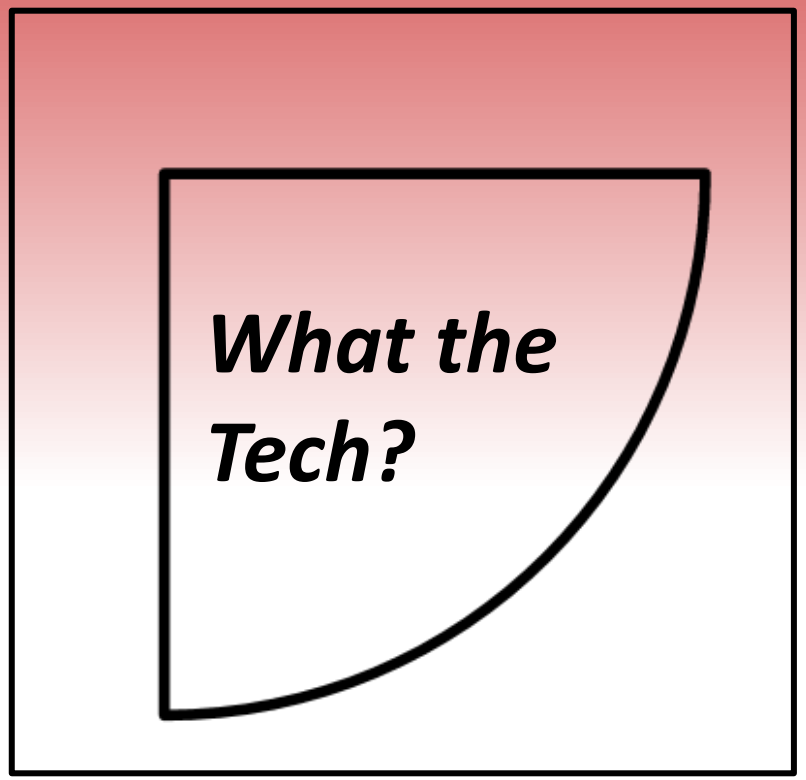
PREDICTIVE ANALYTICS AND AI



Last time, we…

- Shared our third round of weekly reflections
- Did a deep dive on Predictive Analytics & AI
  - KNIME Activity

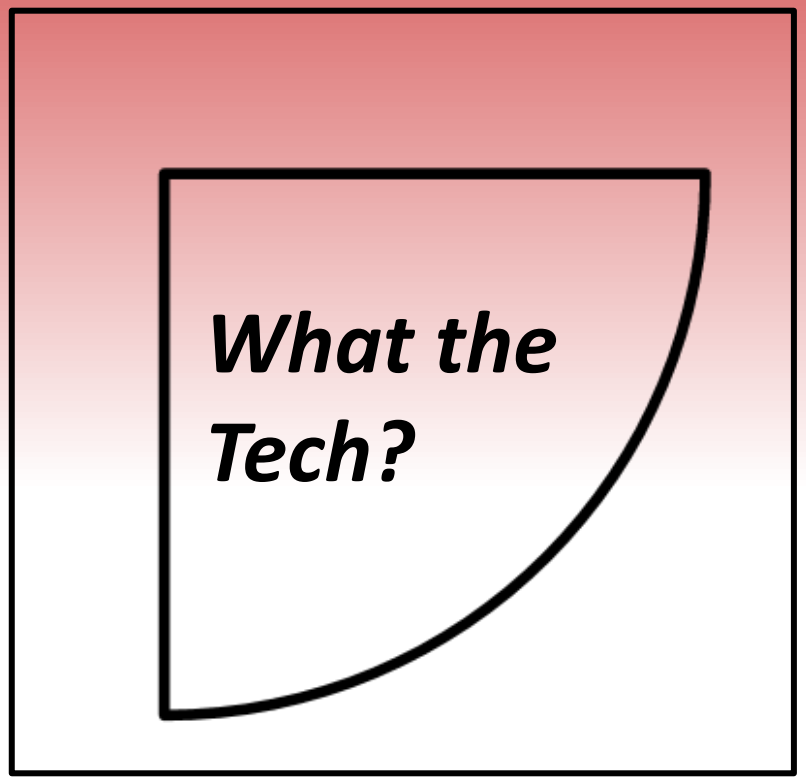
PREDICTIVE ANALYTICS AND AI



Today, we will:

- Continue Predictive AI discussion
  - Terms
  - Group activity: Teachable Machine (Google)
  - Ethics
- Talk about the final project, etc.
- Debrief
- Look ahead at Week 5
  - Generative AI

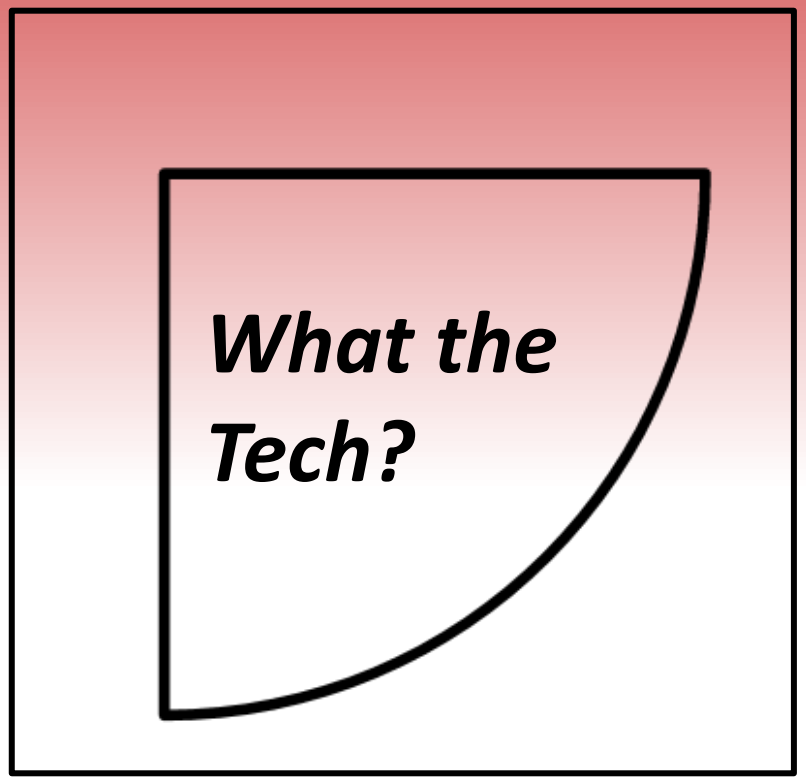
PREDICTIVE ANALYTICS AND AI



Terms used last time

- We heard some new terms:
  - **Normalization**
  - **K-Means**
  - **Decision Trees**
  - **Random Forest**
- Mostly related to programming side of AI

DEFINITION OF TERMS

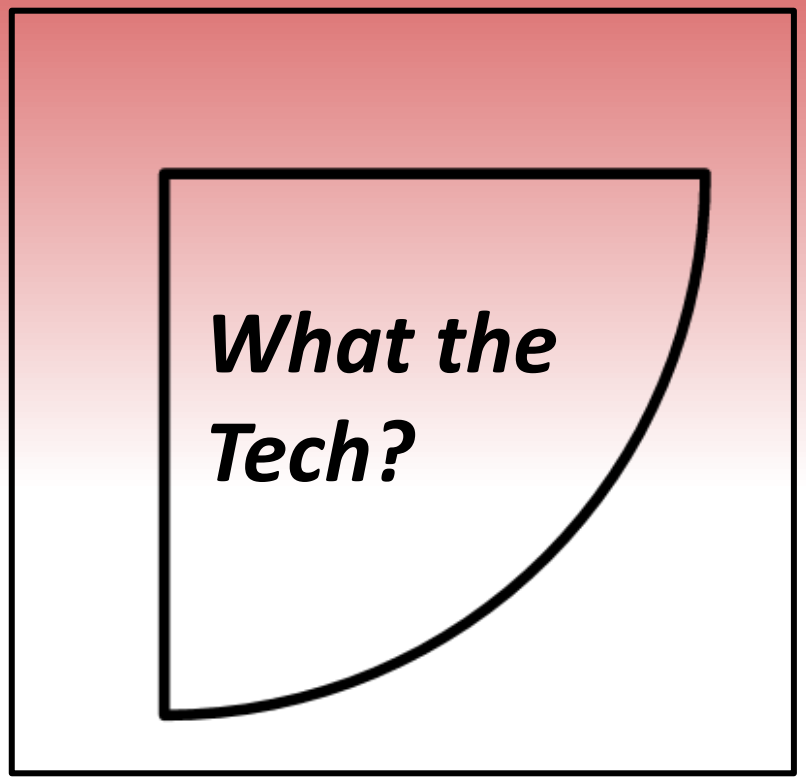


We will be explaining some of the terms used last time

- **Normalization:** Normalization is the process of adjusting values to a common scale to enable fair comparisons and analysis.

- **K-Means:** K-means is a clustering machine learning algorithm that organizes data points into clusters based on how close they are to each other, with the aim of making each cluster as compact and distinct from the others as possible.

DEFINITION OF TERMS

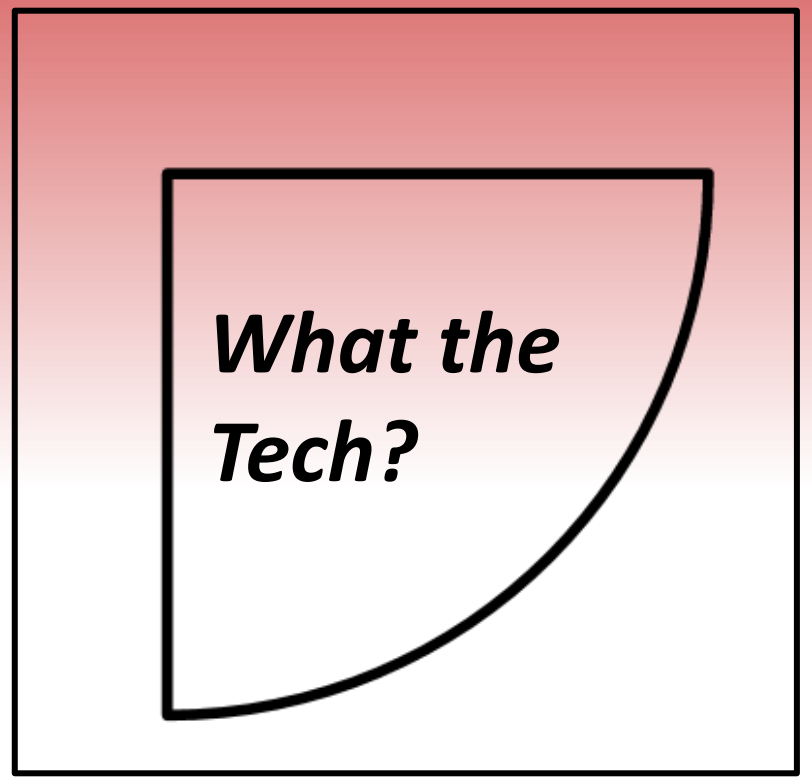


We will be explaining some of the terms used on last time

- **Decision Trees:** A decision tree is a machine learning model that represents decisions and their potential consequences in a tree-like structure. This structure consists of nodes representing decision points, branches representing possible outcomes of those decisions, and leaf nodes representing the final outcome or prediction.

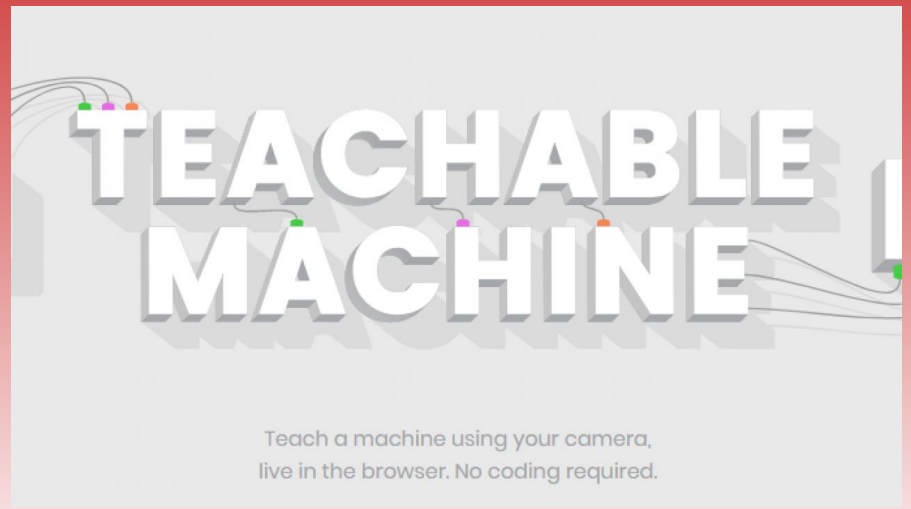
- **Random Forest:** A random forest model uses the prediction from multiple decision trees and combines their predictions to improve accuracy and robustness of the final decision.

PREDICTIVE ANALYTICS AND AI



Any questions? Refreshers?

Hands-On Activity: Teachable Machine



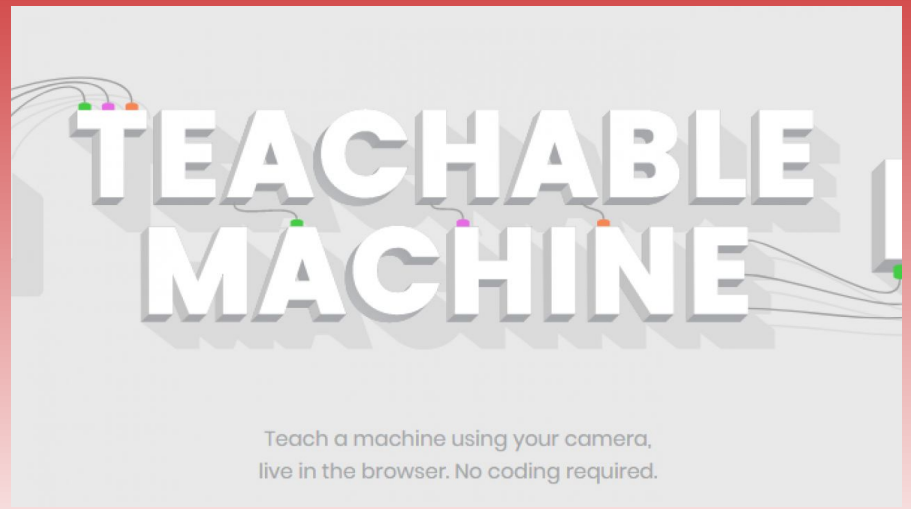
Teachable Machine

- Developed by Google.
- A web-based program that makes "machine learning models."
- Marketed as being easy to use.

Who has used Teachable Machine before?

- *"You can currently train Teachable Machine with images (pulled from your webcam or image files), sounds (in one-second snippets from your mic), and poses (where the computer guesses the position of your arms, legs, etc from an image)."* -From the FAQ

Hands-On Activity: Teachable Machine



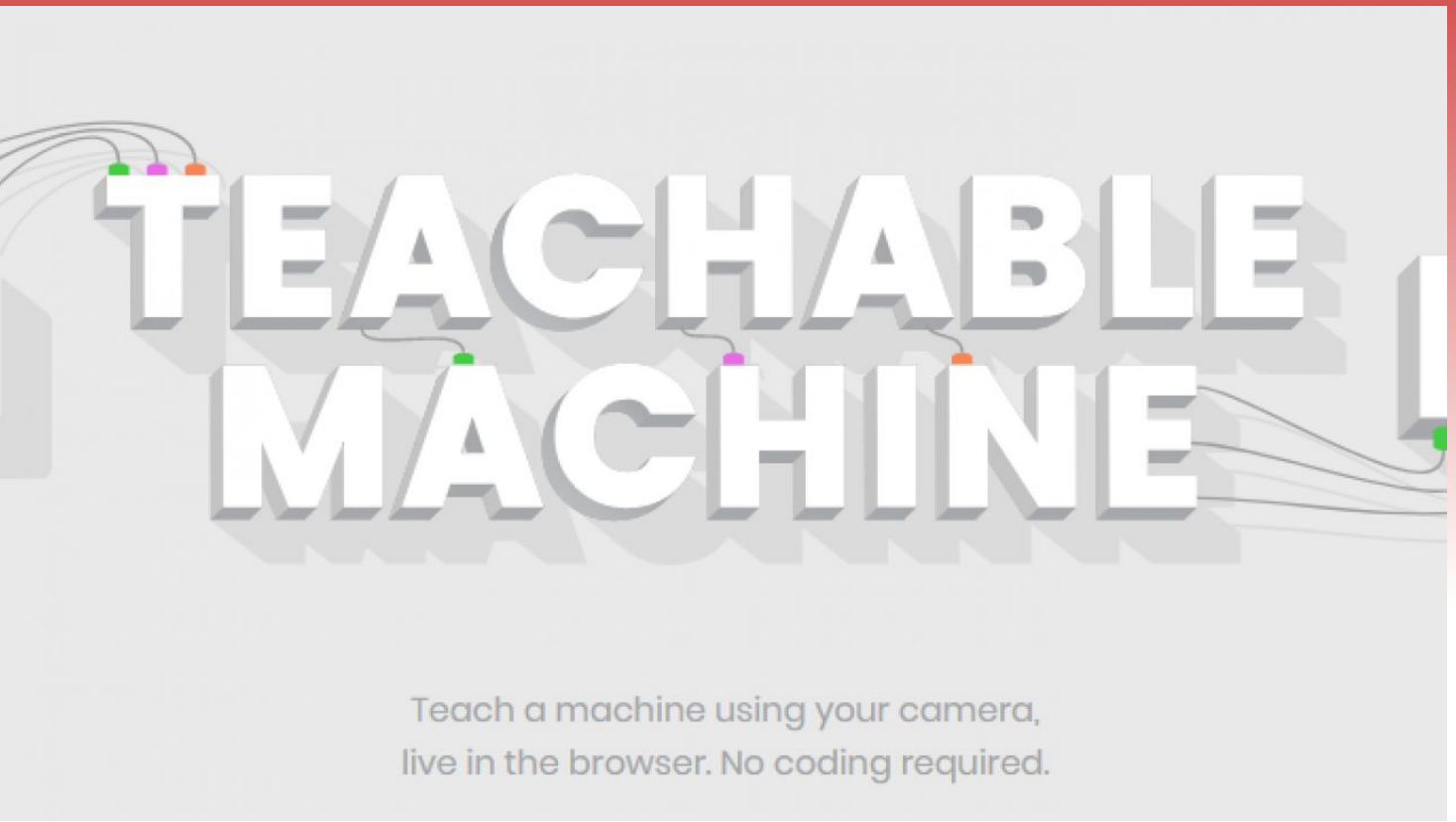
Teachable Machine

- Developed by Google.
- A web-based program that makes "machine learning models."
- Marketed as being easy to use.

Who has used Teachable Machine before?

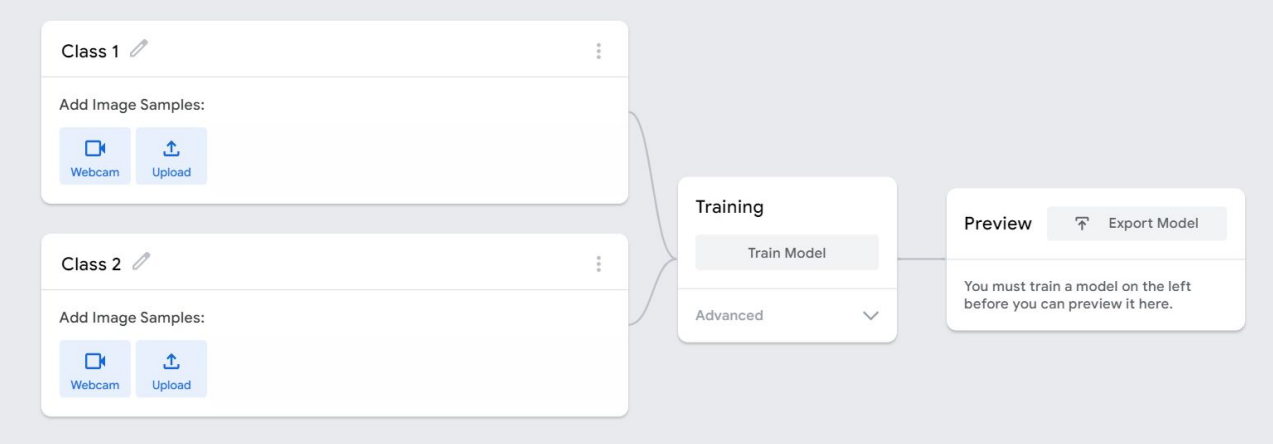
- *"You can currently train Teachable Machine with images (pulled from your webcam or image files), sounds (in one-second snippets from your mic), and poses (where the computer guesses the position of your arms, legs, etc from an image)."* -From the FAQ

Hands-On Activity: Teachable Machine



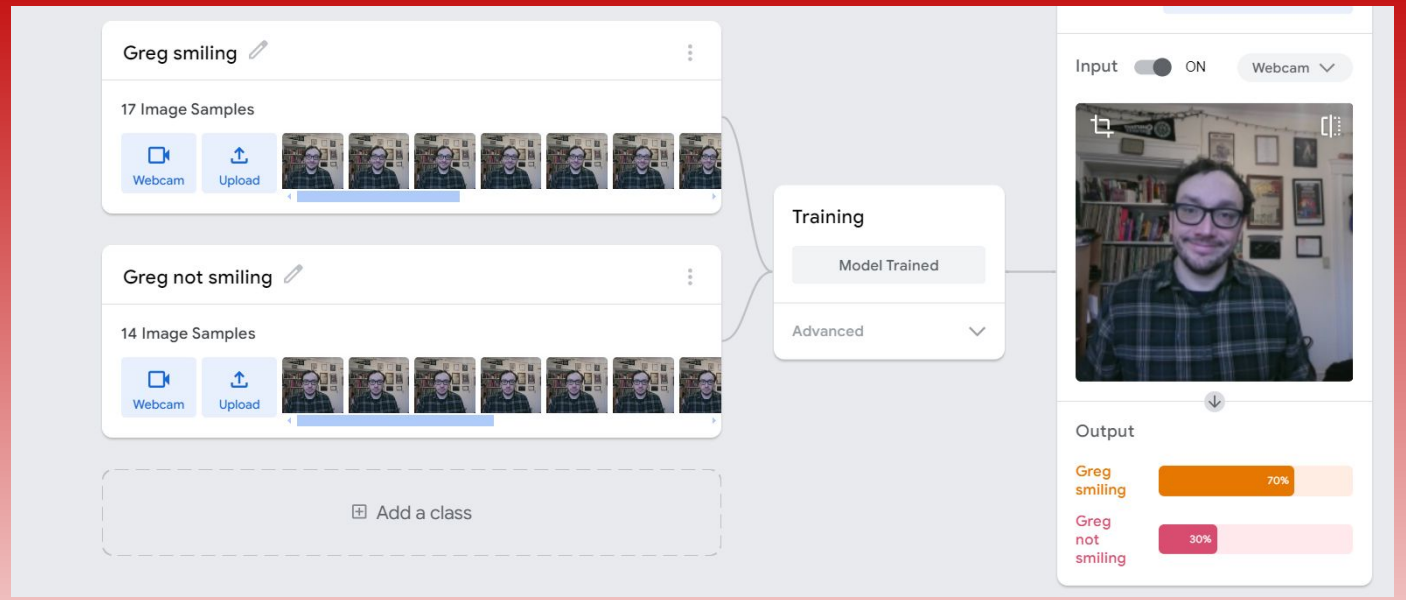
- If you feel like joining in, go to https://teachablemachine.withgoogle.com/
- Don’t feel comfortable? No sweat! Remember, we’re here to have fun.
- Ask yourself as we go: how could this data be beneficial? Or not?

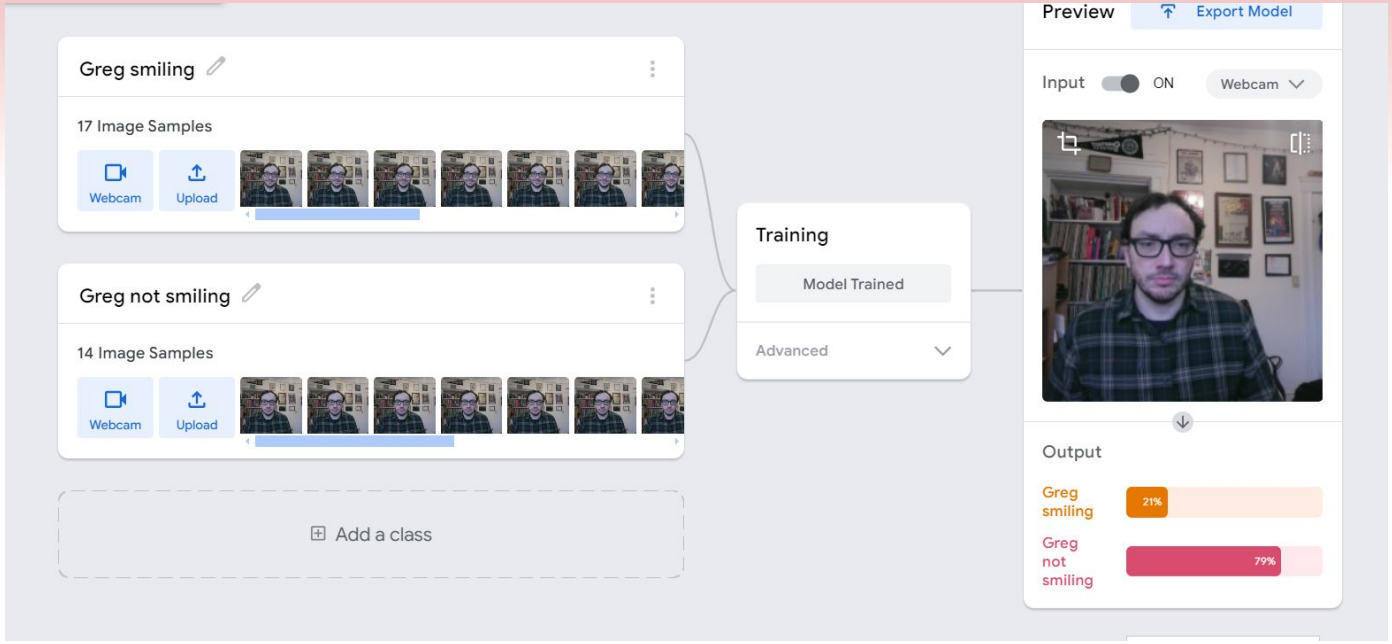
Hands-On Activity: Teachable Machine



Steps

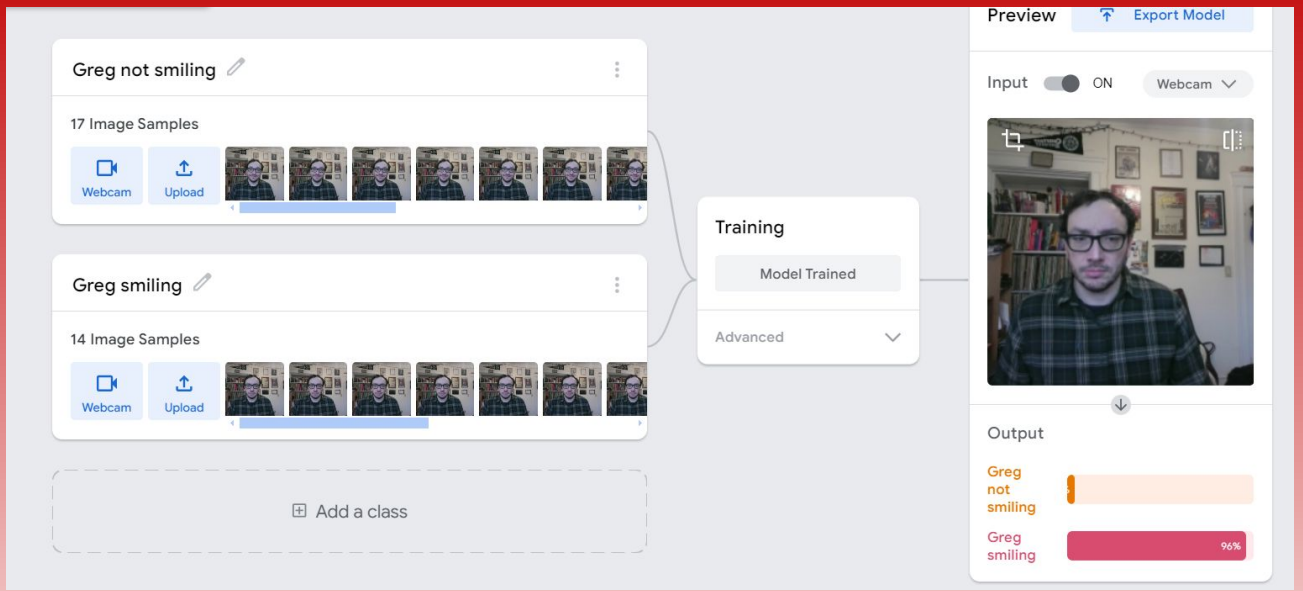
1. Label “Class One” as “Smiling”
2. Label “Class Two” as “Not Smiling”
3. Click on the Webcam button for the Smiling class and smile as you take a few pictures
   a. You may have to grant the program permission
4. Do the same for “Not Smiling”
5. Click “Train Model”
6. See what happens…

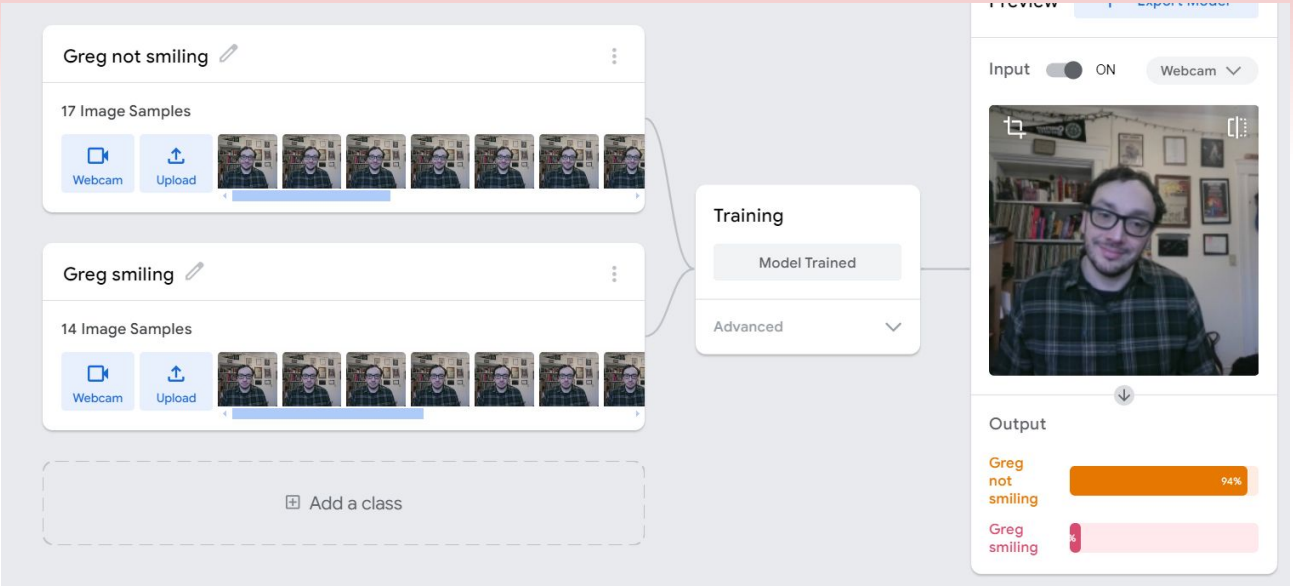




Hands-On Activity: Teachable Machine

- Your results should look something like this.
- Congrats! You've trained a model to identify whether or not you're smiling with fairly high accuracy.
- But what happens if we change the classifications…
  - Change Class One to **Not Smiling**, Class Two to **Smiling**, and click "Train Model"...



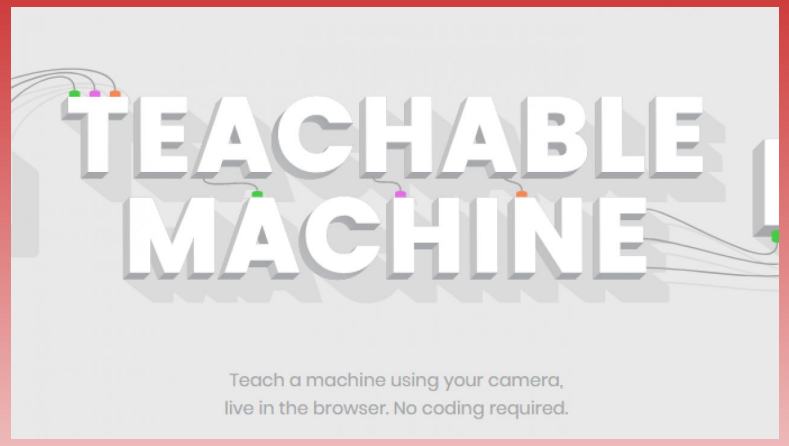


Hands-On Activity: Teachable Machine

- Notice anything different?
- The model we trained says that it can still tell the difference between smiling and not smiling.

Hands-On Activity: Teachable Machine

What do we know now?



Lots of Applications

- Risk assessments
- Strategy development
- Drafting policy
  - Planning for all these in the long term.
  - This lets AI do things we can't do (or at least quickly).

Lots of Limits

- Knowledge limit/lack of accuracy
- Unforeseen circumstances

Lots of Risks

- Ethical risks
- Practical risks
- What happens if the data are flawed?

Let's look at a few examples!

EXAMPLE 1

Predictive Algorithm:

This is what the HR department's budget should be in the new fiscal year

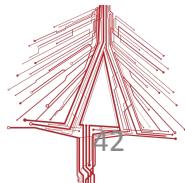
What could a limit be?

Predictive Input

Company's hiring/firing rates

Predictive Output:

Give HR extra $10k to improve screening & training



EXAMPLE 2

Predictive Algorithm:

These are the players we need to sign by 2030

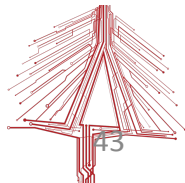
What could a limit be?

Predictive Input

A team's win-loss ratio

Predictive Output:

Four new starting pitchers



EXAMPLE 3

Predictive Algorithm:

These are the optimal release dates for a new album by genre and geography.

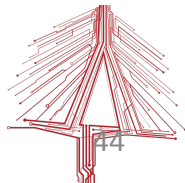
What could a limit be?

Predictive Input

Most streams of a new song by region

Predictive Output:

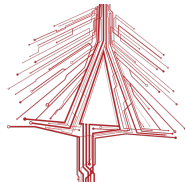
Move Dec. 2024 cumbia releases in South America to Oct.



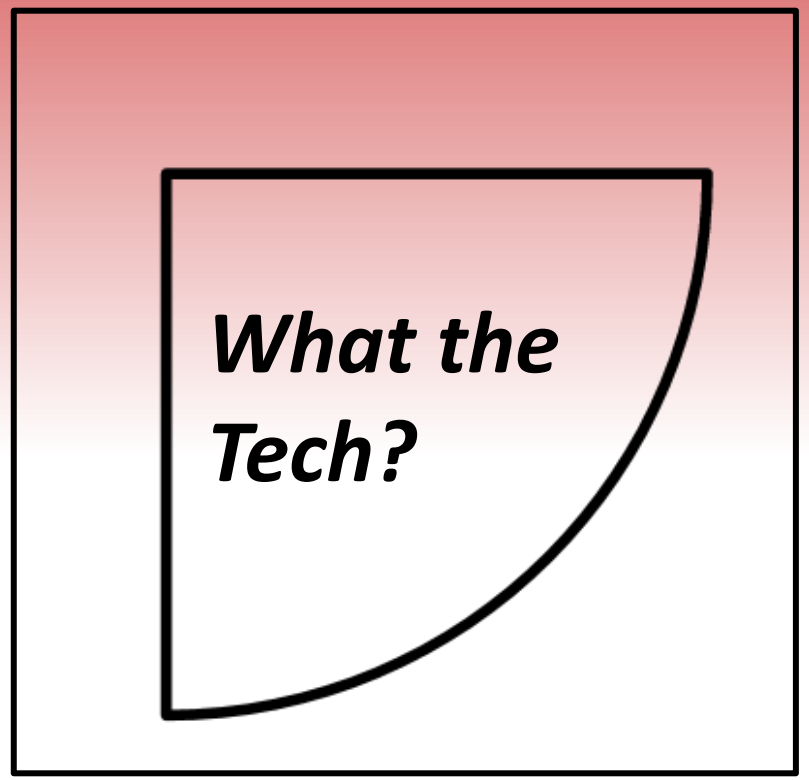
Questions so far?

Break time!

Next: Predictive AI and Ethics



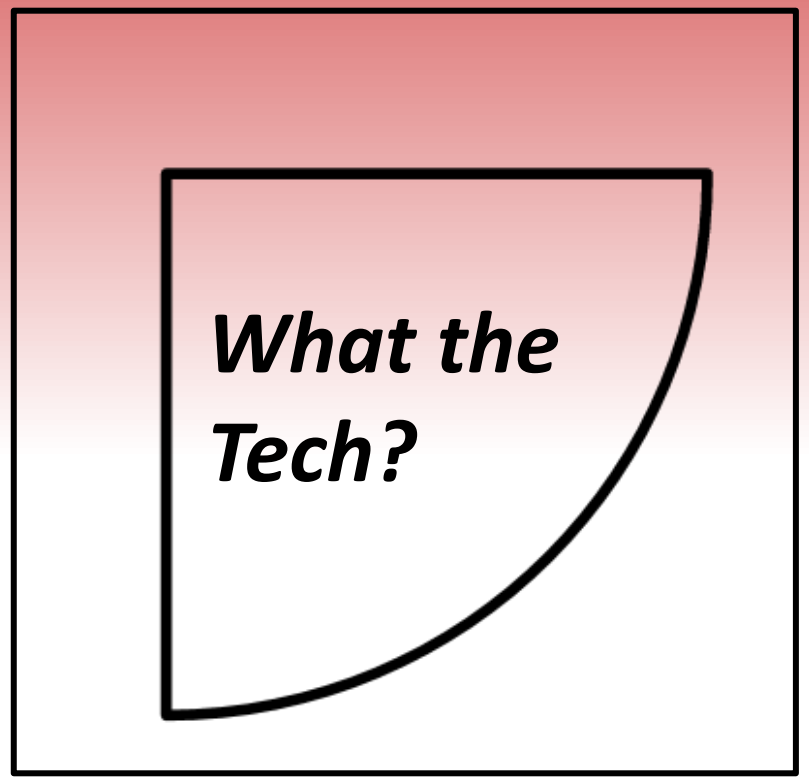
PREDICTIVE ANALYTICS AND AI



The Ethics of Predictive Analytics and AI

- "Predictive AI" has many positive uses
- But even the most accurate data is a liability when it's misapplied.
- When we make mistakes with our predictions, what are the consequences and for whom?

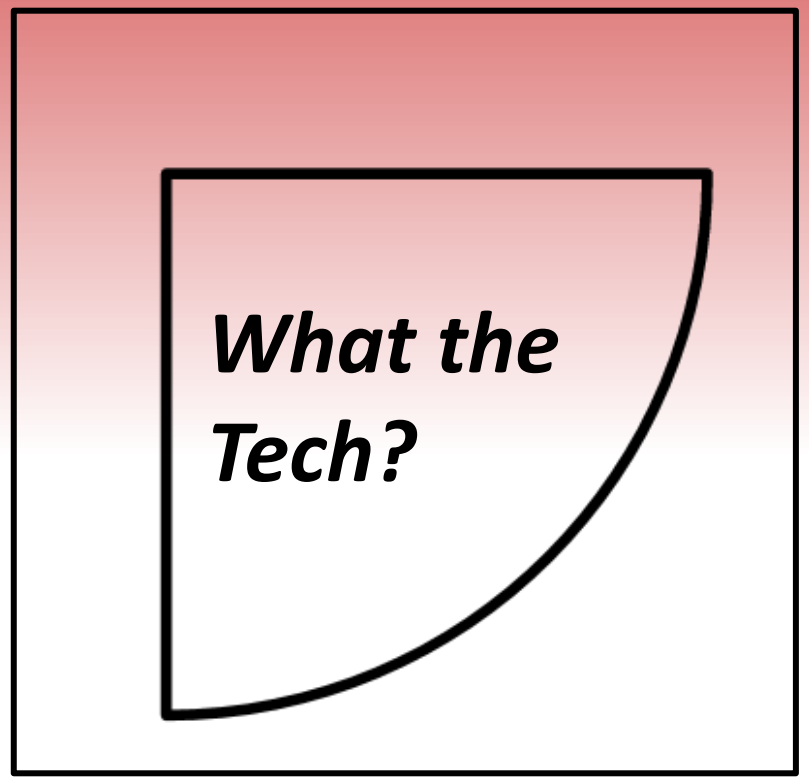
PREDICTIVE ANALYTICS AND AI



The Ethics of Predictive Analytics and AI

- Let's look a few types of ethics:
  - Unequal access
  - Hacking and privacy
  - Unanticipated impacts

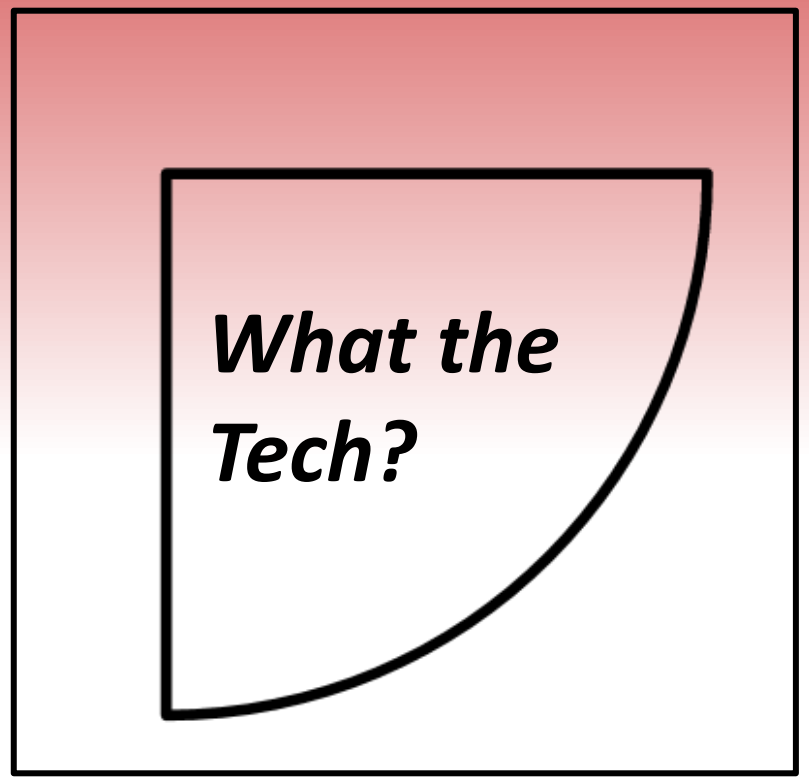
PREDICTIVE ANALYTICS AND AI



Unequal Access

- **Sports:** Since *Moneyball,* teams have used predictive analytics to develop better lineups and strategies.
  - But what happens when wealthy teams can do the same but with deeper pockets?
  - Are these strategies always in the best interests of the players?
- **Municipalities**: Do some cities and towns have more resources to build predictive models for action?
- **Communities:** Are community members included in conversations about what we should use predictive AI to do and how we do it?

PREDICTIVE ANALYTICS AND AI



Hacking and Privacy

- **Data usage:** How is the data collected? Did I give consent?
- **Access:** Are any third parties looking at the data?
- **Retention:** How long will this data be held for?
  - Policies vary (or may not even exist!)

UNANTICIPATED IMPACTS

- Problem: a string of fires in condemned properties in Jamaica Plain from 2017-2023
- Solution: create a database of condemned, fire-prone properties by 2024.
- What are the consequences of being prepared for a fire that doesn't happen?

- Problem: recurrence of robbery condemned properties in Jamaica Plain from 2017-2023
- Solution: increased police patrols and preventative detention around those properties
- What are the consequences of arresting a person who wasn't going to commit a crime?

Predictive Algorithm:
Based on the patterns the machine learned from the input:

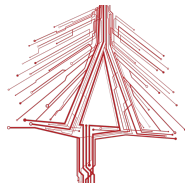
These are how many officers the police department should deploy

Predictive Input
Examples of data/information:

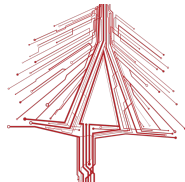
A particular type of recurring crime

Predictive Output:
Predictions or actions responding to the input:

Sending “X” amount of officers to this neighborhood at this time

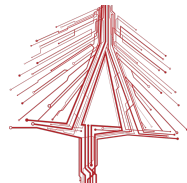


Questions so far?

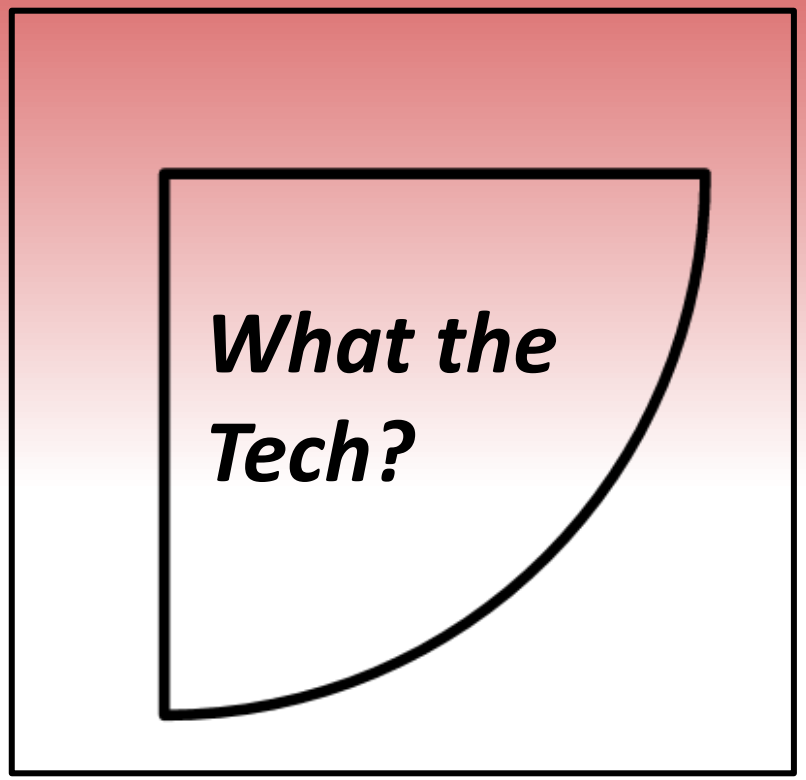


FINAL PROJECT

1) Break into groups (no more than three people) based on common community type (remember week one)
2) Pick a situation/use case related to that community where AI could play a part.
3) Define potential ethical limits and issues
4) Present findings via a media form of your choosing (video, slideshow, etc.)
5) In addition to your media presentation, write two paragraphs that summarize your project and explain how this issue also impacts your city, whether as a whole or in your respective neighborhood(s).



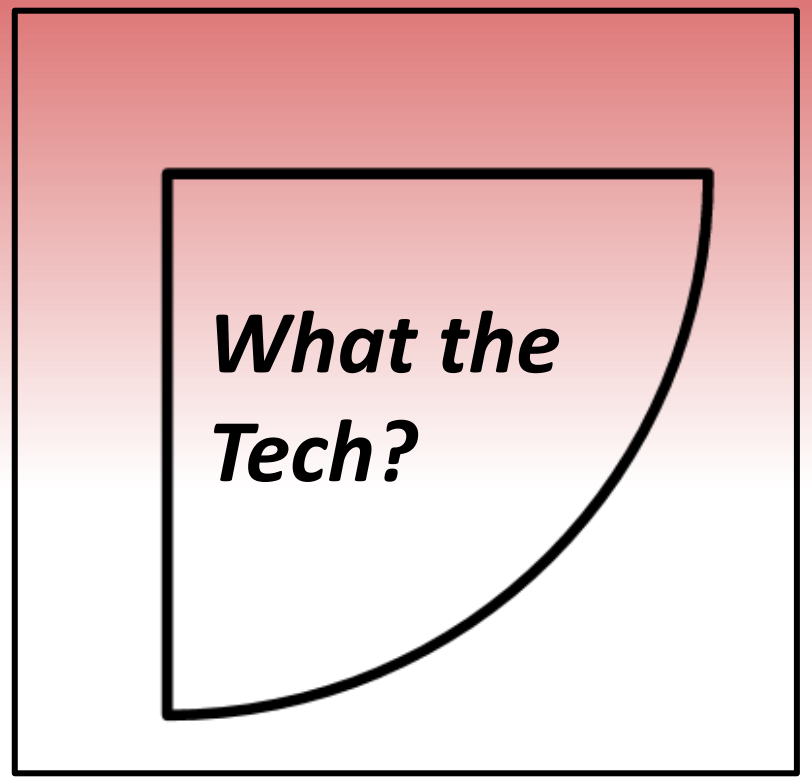
WTT FINAL PROJECT EXAMPLE



The group's write-up ("memo") included:

- Background
  - Their community
  - AI and their community
- Options
  - Positive impacts
  - Potential ethical issues
- Recommendations
  - Important values for AI policies
  - Suggestions

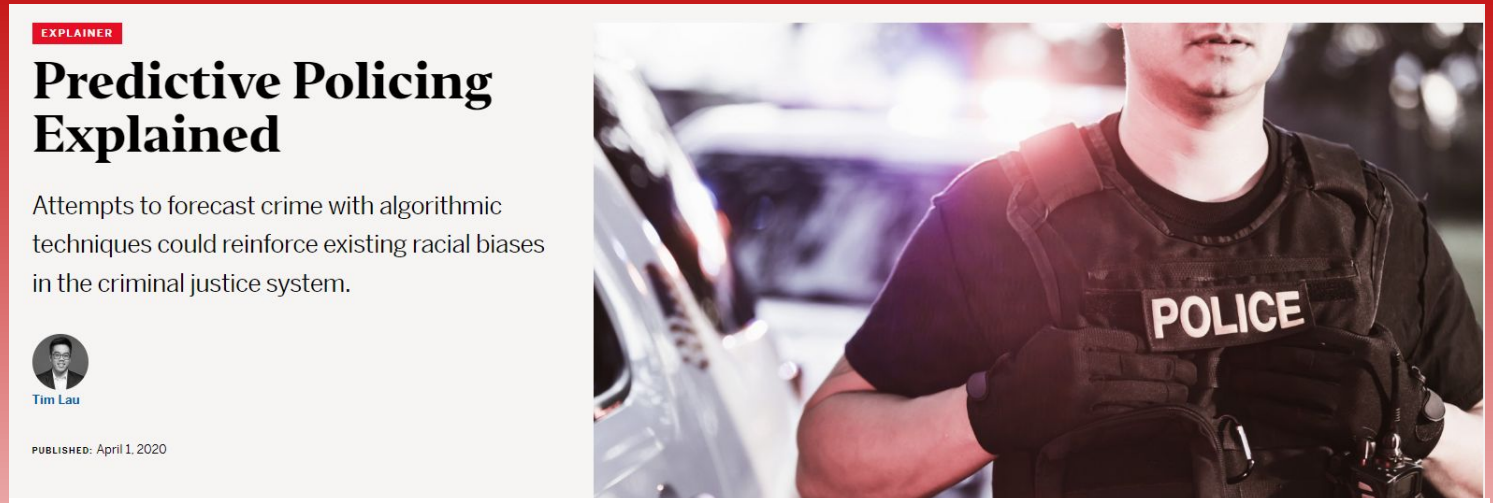
PREDICTIVE ANALYTICS AND AI



- Debrief!
- A chance for you to tell us how it's going. For instance:
  - What went well this week?
  - What didn't?
  - What are you excited for?
  - What are you unsure about?

PREDICTIVE ANALYTICS & AI: BONUS READING

- From April 2020
- Concise guide to predictive analytics, AI, and policing
- Highlights include
  - Definition & examples
  - Legal and ethical concerns
  - The data used



EXPLAINER

Predictive Policing Explained

Attempts to forecast crime with algorithmic techniques could reinforce existing racial biases in the criminal justice system.

Tim Lau

PUBLISHED: April 1, 2020

BRENNAN CENTER FOR JUSTICE

https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained

END OF CONTENT