Session 1 Handout: How Predictions Get Made

FAST-05 Applied Data Science and ML for Business Users · 7formations · Instructor: Wadood Chaudhary · All examples fictional

Why this session matters

You already work with predictions: lead scores in the CRM, reorder suggestions in the ERP, fraud or quality flags, staffing forecasts. You don't need to build models to use them well. You need to know what goes in, what comes out, and who decides. This session gives you that vocabulary and a one-page prediction map.

Key concepts

1. Rules vs. learned patterns

A **rule** is written by a person: "Flag claims over $10,000." "Reorder below 200 units." It's transparent and does exactly what it says.

A **model** learns patterns from many past examples: "Claims filed soon after a policy starts, at night, with no police report have tended to be fraud." It can find patterns people miss, but it's only as good as the history it learned from, and it can be confidently wrong when the world changes.

2. Training data, features, labels

Term	Plain meaning	Clinic no-show example
Training data	Past examples the model learns from	2023-2024 appointments
Features	Inputs known at prediction time	Lead time, weekday, reminder sent, prior misses
Label	The past outcome to learn	Attended or no-show
Prediction	The model's output for a new case	Risk score 0.62 for next Tuesday 9 AM

Supervised learning = learning from examples where the label is known.

3. Confidence scores

Most business models output a **score** (often 0 to 1). A score **ranks** cases by how much they resemble past positives. A score of 0.55 is not "55% likely" unless the model has been specifically calibrated, and it's never a decision. Someone chooses a **threshold** ("call patients above 0.6") and someone **acts**. Those are business choices.

4. Data rules for analysis tools

Copilot in Excel, Power BI Copilot, ChatGPT data analysis, and Gemini in Sheets make it easy to upload whole files. Before you do:

- use only approved tools with work accounts;
- remove columns you don't need (names, IDs, contact info);
- never upload personal information (G.L. c. 93H, 201 CMR 17.00) or PHI unless the tool is approved for it under a business associate agreement;

- "de-identified" has a specific legal meaning under HIPAA. Removing names alone is not enough.

Worked example: prediction map

Field	Charles River Family Dental no-show score
Predicts	Whether a booked appointment will be missed
Inputs	Lead time, weekday, reminder sent, prior misses
Learned from	2023-2024 appointment outcomes
Output	Score 0-1 shown on the schedule
Who acts	Front desk calls appointments above 0.6 the day before
Wrongly says yes	An extra call to a reliable patient (minor annoyance)
Wrongly says no	Empty chair: lost hygienist hour (~$150, fictional)
Data notes	IDs only in analysis; no diagnosis or health data

Checklist: understanding a prediction

- ☐ I can say what it predicts and from what inputs.
- ☐ I know what past outcome it learned from, and how old that data is.
- ☐ I know who acts and what they do.
- ☐ I can state the cost of each error direction.
- ☐ I know what data I may and may not put into analysis tools.

Do and don't

Do	Don't
Ask "who acts on this score?"	Treat a score as a decision
Describe features in plain words	Assume the model "understands" your business
Check how old the training data is	Assume past patterns still hold
Strip unneeded columns before analysis	Upload whole customer files "for context"
Treat a score as a ranking	Read 0.55 as "55% chance" without checking

Try this at work

1. List every score, flag, or forecast you see in your systems.
2. Pick one and complete a prediction map.
3. Ask its owner one question you couldn't answer.

Glossary

Term	Meaning
Model	A pattern learned from past data that makes predictions
Supervised learning	Learning from examples with known outcomes
Feature	An input to the model
Label	The known outcome in training data
Score	Model output ranking how much a case resembles past positives
Threshold	The score above which the business takes action

Key takeaways

- Models learn patterns from past examples; rules are written by people.
- Scores rank; thresholds and people decide.
- Map every prediction: inputs, output, who acts, cost of errors.