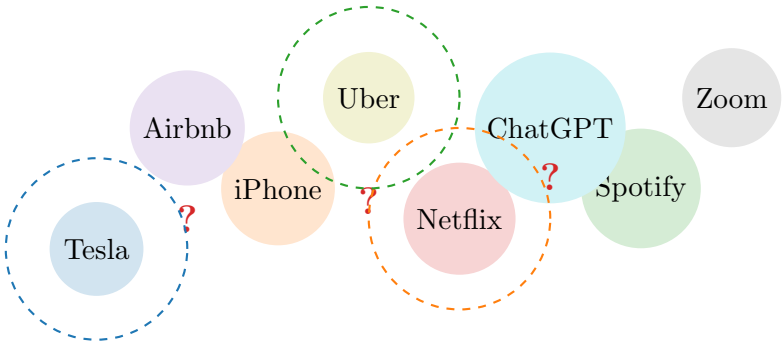
Machine Learning for Innovation

A Discovery Journey with Real Data

Challenge

Can You Beat the Algorithm?

Inside: 8 challenges where human intuition fails
and machine learning reveals hidden innovation patterns



Name: _____

Date: _____

Your Field: _____

Did You Know?

Companies using ML for innovation pattern discovery are **2.3x more likely** to be industry leaders

Challenge 1: Group These Real Innovations

Exercise: Beat the Algorithm - Round 1

Below are 20 real innovations. Group them into 4 categories. You have 3 minutes!

Tesla Model 3 Electric Vehicle 2017	**iPhone** Smartphone 2007	**Netflix** Streaming 1997	**Airbnb** Sharing Economy 2008
Spotify Music Streaming 2006	**Uber** Ride Sharing 2009	**ChatGPT** AI Assistant 2022	**Zoom** Video Conferencing 2011
Instagram Social Media 2010	**PayPal** Digital Payment 1998	**Amazon Prime** Subscription 2005	**Google Maps** Navigation 2005
WhatsApp Messaging 2009	**Bitcoin** Cryptocurrency 2009	**Slack** Team Communication 2013	**TikTok** Short Video 2016
Peloton Fitness Tech 2012	**Discord** Gaming Chat 2015	**DoorDash** Food Delivery 2013	**Robinhood** Trading App 2013

Your 4 Groups

Group 1: _____
Group 2: _____
Group 3: _____
Group 4: _____

Compare with Your Peer

Ask someone else how they grouped them. Are they the same?
Differences found: _____
Who's right? _____

Challenge

The Reveal: ML found 7 different valid groupings using different features!

- By founding year: Pre-2010 vs Post-2010 vs Recent
- By business model: Platform vs Product vs Service
- By target: B2C vs B2B vs Both
- By disruption level: Industry creators vs Enhancers

Key Insight: There's no single "correct" grouping - it depends on which features matter for your goal!

Challenge 2: The Spotify Prediction Failure

Real Case Study

Spotify processes 60,000 new songs daily. How do they know which songs you'll like?

Exercise: Predict the Playlist

Here are 10 real songs with 5 Spotify features. Circle the 3 songs that would appear in the same playlist:

Song	BPM	Energy	Dance	Acoustic	Valence
"Blinding Lights" - Weeknd	171	0.73	0.67	0.00	0.39
"Shape of You" - Ed Sheeran	96	0.65	0.83	0.58	0.93
"Bohemian Rhapsody" - Queen	72	0.40	0.29	0.27	0.22
"Old Town Road" - Lil Nas X	136	0.62	0.88	0.03	0.64
"Someone Like You" - Adele	68	0.33	0.60	0.95	0.12
"Levitating" - Dua Lipa	103	0.82	0.70	0.00	0.91
"Stairway to Heaven" - Zeppelin	82	0.35	0.33	0.36	0.19
"WAP" - Cardi B	133	0.84	0.93	0.10	0.35
"Perfect" - Ed Sheeran	95	0.45	0.60	0.16	0.37
"Thunder" - Imagine Dragons	168	0.81	0.60	0.01	0.29

Your 3 songs: _____, _____, _____

The Spotify Reality

Plot Twist! Spotify actually uses **50+ audio features**, not just 5:

- Loudness
- Speechiness
- Instrumentalness
- Liveness
- Tempo variance
- Key signature
- Time signature
- Mode (major/minor)
- Duration
- Popularity trend
- Skip rate
- Completion rate
- Playlist adds
- User demographics
- Time of day patterns
- Seasonal trends
- And 34 more...

Did You Know?

Spotify's Discover Weekly uses clustering on these 50+ features to find songs similar to your taste profile. It generates **2 billion** playlist recommendations every Monday!

Your Discovery

Why your prediction likely failed:
With only 5 features visible, you missed: _____
Key Learning: Human intuition breaks down beyond 3-4 dimensions. ML thrives in 50+ dimensions!

Challenge 3: Smartphone Market Segmentation

Exercise: Group These Real Phones (2024 Models)

A phone manufacturer wants to understand market segments. Group these 15 phones into 3 categories:

Model	Price ($)	Screen (inch)	Battery (mAh)	Camera (MP)	RAM (GB)	5G
iPhone 15 Pro Max	1,199	6.7	4,422	48	8	Yes
Samsung S24 Ultra	1,299	6.8	5,000	200	12	Yes
Google Pixel 8 Pro	999	6.7	5,050	50	12	Yes
OnePlus 12	799	6.82	5,400	50	16	Yes
iPhone 15	799	6.1	3,349	48	6	Yes
Samsung A54	449	6.4	5,000	50	8	Yes
Pixel 8a	499	6.1	4,492	64	8	Yes
Xiaomi 14	649	6.36	4,610	50	12	Yes
Nothing Phone 2	599	6.7	4,700	50	12	Yes
Motorola Edge 40	399	6.55	4,400	50	8	Yes
iPhone SE 3	429	4.7	2,018	12	4	Yes
Samsung A15	199	6.5	5,000	50	4	No
Redmi Note 13	249	6.67	5,000	108	8	No
Nokia G60	299	6.58	4,500	50	6	Yes
Moto G Power	179	6.6	5,000	50	4	No

Your Manual Clustering

Premium Segment: _____

Mid-Range Segment: _____

Budget Segment: _____

Challenge

Now re-cluster using ONLY these features and see how groups change:

1. **By Battery Life Only:** Long-life vs Standard vs Compact
2. **By Camera Only:** Photography-focused vs Standard vs Basic
3. **By Ecosystem:** Apple vs Samsung vs Google vs Others

Discovery: Different features = Different market insights!

Real Case Study

Samsung's Real Strategy: They use clustering on 127 features including user behavior, app usage, purchase history, and demographics to identify micro-segments like "Mobile Gamers," "Photography Enthusiasts," and "Business Power Users."

Challenge 4: Innovation Distance Calculator

Exercise: Calculate Real Innovation Distances

Compare these transportation innovations using actual metrics:

Feature	Uber	Lyft	Traditional Taxi
Average wait time (min)	5	6	12
Price per mile ($)	2.20	2.10	3.50
App rating (1-5)	4.2	4.3	2.8
Driver rating system	Yes (1)	Yes (1)	No (0)
Cashless payment	Yes (1)	Yes (1)	Sometimes (0.5)
Price transparency	Yes (1)	Yes (1)	No (0)

Calculate Euclidean Distance

Between Uber and Lyft:

$$d = \sqrt{(5-6)^2 + (2.20-2.10)^2 + (4.2-4.3)^2 + \ldots} = \_\_\_\_$$

Between Uber and Taxi:

$$d = \sqrt{(5-12)^2 + (2.20-3.50)^2 + (4.2-2.8)^2 + \ldots} = \_\_\_\_$$

Calculate Manhattan Distance

Between Uber and Lyft:

$$d = |5-6| + |2.20-2.10| + |4.2-4.3| + \ldots = \_\_\_\_$$

Between Uber and Taxi:

$$d = |5-12| + |2.20-3.50| + |4.2-2.8| + \ldots = \_\_\_\_$$

Challenge

Which distance metric makes more sense here?
Manhattan distance treats each feature independently (wait time doesn't affect price).
Your answer: _____
Real Insight: That's why Uber and Lyft cluster together - they're innovative in the same dimensions!

Did You Know?

When Uber analyzes competition, they track 200+ metrics including surge pricing patterns, driver availability heat maps, and user switching behavior. Their clustering algorithm identified that their real competition in Manhattan isn't Lyft - it's the subway during rush hour!

Challenge 5: Netflix's Hidden Genres

Real Case Study

Netflix doesn't just have "Comedy" or "Drama." They have 76,897 micro-genres like:

- "Critically Acclaimed Emotional Underdog Movies"
- "Violent Sci-Fi from the 1980s"
- "Sunday Night Crime Shows for Couples"

Exercise: Find the Hidden Pattern

These 20 shows all belong to ONE secret Netflix genre. Can you identify it?

- Breaking Bad
- Better Call Saul
- Ozark
- Narcos
- The Sopranos
- Peaky Blinders
- The Wire
- Boardwalk Empire
- Mad Men
- House of Cards
- Succession
- Billions
- Ray Donovan
- The Americans
- Homeland
- Queen of the South
- Power
- Yellowstone
- Mare of Easttown
- True Detective

Your guess at the genre: _____

The Netflix Answer

Genre: "Dark Antiheroes in Morally Complex Dramas"
Netflix's clustering algorithm found these commonalities:

- Protagonist moral ambiguity score: ¿0.8
- Episode runtime: 45-60 minutes
- Violence level: Medium-High
- Viewer completion rate: ¿75%
- Binge-watching coefficient: ¿0.7
- Male viewership: 60-70%
- Peak viewing: 9-11 PM

Exercise: Calculate Silhouette Score

If these shows form a cluster, calculate how well "The Office" fits:

- Average distance to shows in cluster: 8.5
- Average distance to next nearest cluster (Comedies): 3.2
- Silhouette = $\frac{b-a}{\max(a,b)} = \frac{3.2-8.5}{\max(8.5,3.2)} =$ ----

Interpretation: Negative score means _____

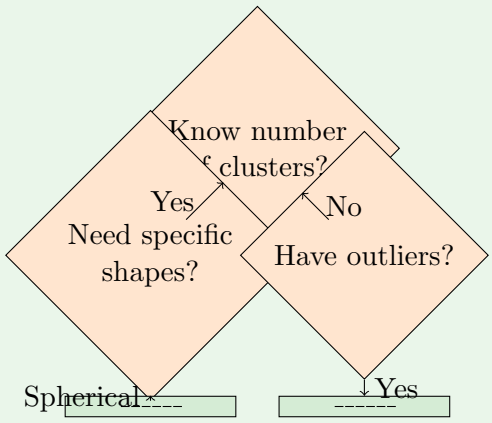
Challenge 6: Which Algorithm for Which Company?

Exercise: Match the Real Scenario to the Right Algorithm

Draw lines connecting each company's challenge to the best clustering algorithm:

Company Challenge	Algorithm
Amazon: "We need exactly 5 customer segments for our marketing campaigns"	**DBSCAN**
Facebook: "Find fake accounts - they're rare and different from normal users"	**K-Means**
Google: "Organize all websites into a hierarchy from general to specific"	**Hierarchical**
Spotify: "Users can belong to multiple music taste groups"	**Gaussian Mixture**
Tesla: "Find defective parts - they cluster in weird shapes on the assembly line"	**Mean Shift**

Decision Flowchart - Fill in the Blanks



Real Case Study

Amazon's Real Implementation:

- Uses K-means on 100+ features for customer segmentation
- Segments: Prime Power Shoppers, Deal Seekers, Brand Loyalists, Window Shoppers, One-time Buyers
- Each segment receives different homepage layouts, email campaigns, and recommendations
- Result: 35% increase in conversion rate

Your Innovation Clustering Project

Exercise: Design Your Own Innovation Analysis

Choose your field and complete this template:

1. Your Innovation Domain

Field: ☐ Tech ☐ Healthcare ☐ Education ☐ Finance ☐ Other: _____

20 Innovations in Your Field:

1. _____ 11. _____
2. _____ 12. _____
3. _____ 13. _____
4. _____ 14. _____
5. _____ 15. _____
6. _____ 16. _____
7. _____ 17. _____
8. _____ 18. _____
9. _____ 19. _____
10. _____ 20. _____

2. Feature Selection

10 Key Features to Track:

1. _____ 6. _____
2. _____ 7. _____
3. _____ 8. _____
4. _____ 9. _____
5. _____ 10. _____

3. Success Metrics

Target number of clusters: _____
Minimum silhouette score: _____
Business question to answer: _____

Your Discovery Reflections

What Patterns Are You Missing?

Before today, I grouped innovations by:

_____

Now I realize I should also consider:

_____

_____

Challenge

The 270-Dimension Challenge
If each innovation has 270 features, and you can only visualize 3 at a time, how many different 3D views would you need to see all possible combinations?
Answer: $\binom{270}{3} = \frac{270!}{3!(270-3)!} = 3,241,350$ views!
Time to view all (at 1 second each): 37.5 days non-stop!
Time for ML to analyze all: ¡ 1 second

Your Top 5 Discoveries

1. _____
2. _____
3. _____
4. _____
5. _____

Real Case Study

Success Story: Procter & Gamble

- Used clustering on 200+ innovation features
- Discovered "Sustainable Millennials" segment
- Launched Tide Eco-Box based on cluster insights
- Result: $100M new revenue stream in year 1

Your Innovation Clustering Action Plan

Week 1 Action Items

☐ Identify 50 innovations in your field
☐ List 20 features that matter
☐ Collect data for at least 10 innovations
☐ Try manual clustering with 3 features
☐ Calculate distances between top 5 innovations
☐ Identify which algorithm fits your needs

Did You Know?

Industry Clustering Applications You Use Daily:

- **Netflix:** 76,897 micro-genres from clustering viewing patterns
- **Spotify:** 5,000 "taste clusters" for Discover Weekly
- **Amazon:** 150 customer segments for personalization
- **Google:** 2 billion web pages organized via clustering
- **LinkedIn:** 147 skill clusters for job matching
- **Instagram:** 32 interest clusters for Explore page

Resources to Explore

Interactive Demos:

- www.tensorflow.org/playground - See clustering in action
- projector.tensorflow.org - Visualize high-dimensional data
- distill.pub/2016/misread-tsne/ - Understand dimensionality

Real Datasets to Try:

- Startup Database: www.crunchbase.com
- Innovation Rankings: www.globalinnovationindex.org
- Patent Database: patents.google.com

Final Challenge

You've discovered that ML can find patterns in 270 dimensions that humans can't see.

Question: What innovation opportunities might be hiding in YOUR data?

Next week: We'll use these techniques on real innovation datasets!

Course Instructors — ML for Innovation — Week 1 Complete