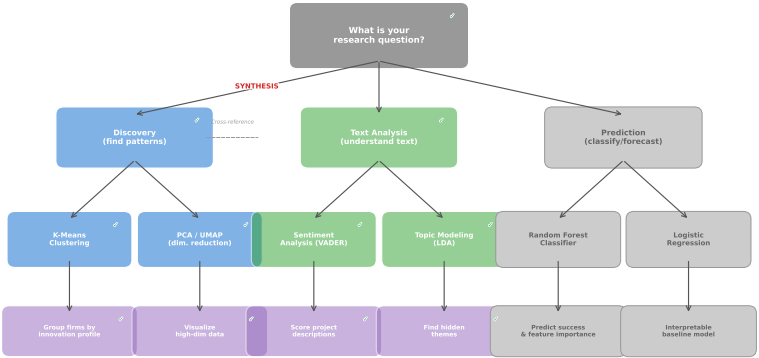
ML Technique Decision Tree



Your research question determines your method. We are cross-referencing the Discovery and Text branches.

Combining Discovery and Text

The power of integration:

We clustered projects by *quantitative* features. We analyzed their *text* with NLP.

Now combine:

- Do different clusters use different language?
- Are some topics concentrated in specific clusters?
- Does sentiment vary across innovator types?

Multi-method research design:

1. **Phase 1:** Segment with clustering
2. **Phase 2:** Profile segments with NLP
3. **Phase 3:** Test differences statistically

Thesis angle

"A mixed-methods ML approach to identifying and characterizing innovation archetypes."

Two branches, one dataset. When independent methods converge on the same story, your findings become defensible.

Cross-Referencing the Branches

	Cluster 1 Cautious	**Cluster 2** Digital	**Cluster 3** Radical	**Cluster 4** Collaborative
Topic: Technology	15%	**45%**	25%	20%
Topic: Healthcare	20%	15%	30%	**35%**
Topic: Efficiency	**40%**	25%	10%	15%
Topic: Innovation	10%	10%	**30%**	20%
Topic: Sustainability	15%	5%	5%	10%
Avg Success	48%	72%	68%	61%

Clusters × topics. Two methods, one dataset. Laying them side by side: does the story hold?

Multi-Method Research Design

Pattern 1: Discover then Predict

1. Cluster to find segments
2. Use cluster labels as features
3. Predict outcomes with classification

Pattern 2: Quantify then Explore

1. Extract NLP features from text
2. Combine with survey variables
3. Cluster on the combined feature set

Pattern 3: Validate Across Methods

1. Find clusters in numeric data
2. Find topics in text data
3. Check alignment (convergent validity)

For your DBA thesis

One method is a finding. Multiple methods converging on the same story is evidence. That is triangulation.

Using clustering AND NLP AND cross-tabulation is triangulation — your findings become defensible.

The Map So Far: What We Know

Technique	What We Found	Evidence
PCA / UMAP	Structure exists in the data	2D projections
K-Means	4 innovator archetypes	Cluster profiles
NLP / Sentiment	Language reveals confidence	Sentiment scores
Topic Modeling	5 hidden innovation themes	Topic distributions

But can we predict who will succeed?

Discovery and Text describe what is. Prediction asks what will be.

We have archetypes. We have language patterns. We have cross-references. But prediction requires a different branch.

Your Turn: Where Does This Fit?

"What surprised you about the patterns we found?
Did anything contradict your expectations?"

Take 60 seconds. Discuss with your neighbor.

What's Next: The Prediction Branch

The third branch of the tree:

- **Classification:** Predict success/failure
- **Feature importance:** What drives success?
- **Generative AI:** LLMs for research
- **Structured output:** Auto-code qualitative data
- **Complete map:** Your ML decision framework

Think about:

- What patterns did you notice today?
- Which clusters surprised you?
- What research questions emerged?
- How might these apply to *your* topic?

Discovery and Text branches are complete. Session 2 opens the Prediction branch.

Map Check: Where Are We?

Branches explored:

- ✓ Discovery: 4 innovator archetypes identified
- ✓ Text: language patterns and 5 topics extracted
- ✓ Cross-reference confirms the story holds

Next on the map:

- Prediction branch: can we predict who will succeed?
- Classification, feature importance, GenAI (Session 2)

Session 1 complete. Two branches explored, cross-referenced, validated. Session 2 tackles prediction.