

## ML Technique Decision Tree



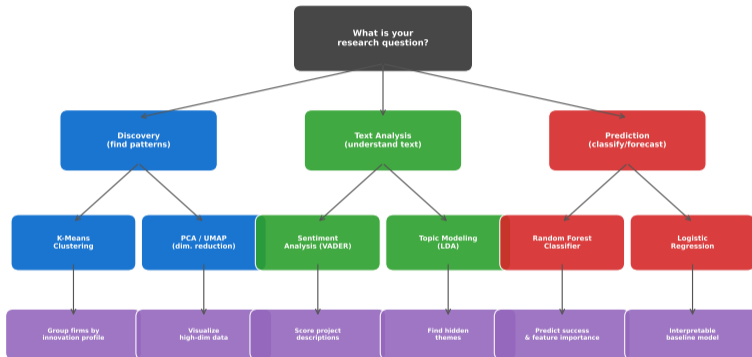
Your research question determines your method. You've walked every branch.

## 7 Branches Explored

#	Technique	Answers	Output
1	PCA / UMAP	Can we visualize structure?	2D projections
2	K-Means Clustering	What groups exist?	Cluster profiles
3	Sentiment Analysis	What tone is used?	Scores (-1 to +1)
4	Topic Modeling (LDA)	What themes emerge?	Topic distributions
5	Classification (RF)	Can we predict outcomes?	Accuracy, importance
6	Logistic Regression	Which factors matter?	Coefficients, p-values
7	Structured Output	Can AI code data?	Structured JSON

Blue = Discovery. Orange = Prediction. Green = Generation. Your research question tells you which branch to follow.

## ML Technique Decision Tree



This is the map you have walked. Now you own it. Start with your research question — the question determines the method.

## Recommended thesis patterns:

1. **Exploratory:**  
Cluster → Profile with NLP → Validate differences
2. **Predictive:**  
Feature engineering → RF + Logistic → Feature importance
3. **Mixed-methods:**  
NLP extraction → Combine with survey → Predict outcomes
4. **Text-first:**  
Topic modeling → Structured output coding → Statistical analysis

## Quality checklist:

- Research question drives method choice
- Proper train/test split or CV
- Multiple methods compared
- Feature importance interpreted
- Limitations honestly discussed
- Results connected to theory
- Code and data available
- AI use disclosed

---

Your thesis could: cluster innovators, profile them with NLP, predict success, then validate with GenAI-coded interviews.

## The Innovation Detective: Evidence Board



**8 modules. 7 techniques. Three branches. From first question to complete toolkit — the map is yours.**

# Which Path Is Yours?

## Research Journey Overview



### Think about:

- What data do you have (or could collect)?
- Is your question about discovery or prediction?
- Do you have text data?

### Share with the group:

- Your research topic (1 sentence)
- Which technique interests you most
- What data challenge you face

---

You have seen all seven tools. Now: which path through the tree is yours?

*“You’ve walked every branch. Which technique would you use first for YOUR thesis, and why?”*

Take 60 seconds. Discuss with your neighbor.

## Learning Python for research:

- *Python for Data Analysis* (McKinney)
- Kaggle free courses ([kaggle.com/learn](https://www.kaggle.com/learn))
- Google Colab — free Jupyter in browser
- ChatGPT/Claude as coding tutor

## ML for social scientists:

- *An Introduction to Statistical Learning* (James et al.)
- *Prediction, Learning, and Games* (Cesa-Bianchi & Lugosi)

## Key papers on ML in innovation:

- Arts et al. (2021): NLP for patent analysis
- Behrens et al. (2020): ML for innovation prediction
- Fernández-Delgado et al. (2014): Classifier comparison

## Tools used today:

- scikit-learn, NLTK, Gensim
- sentence-transformers, UMAP
- See `requirements.txt`

---

The map keeps growing. Full reading list in the course repository.

## Research Journey Overview



---

**From data exploration to structured output — your complete ML research toolkit.**

# The Complete Map

You started with a question.  
Now you have the tools to answer it.  
The next research question is yours.

All materials: [course repository](#)  
Contact: [joerg.osterrieder@uni.ch](mailto:joerg.osterrieder@uni.ch)