The Agony and the...

Master Thesis Proposal

Tristan Leiter

Vienna University of Business and Economics

January 20, 2026

Agenda

Introduction

- **Context:** Modern distributed systems generate massive logs.
- **Pain Point:** Manual analysis is error-prone and slow.
- **Goal:** Develop a hybrid deep-learning model for real-time detection.

Research Gap

Existing solutions struggle with high-dimensional, unlabeled data streams.

Research Question

- **Context:** Modern distributed systems generate massive logs.
- **Pain Point:** Manual analysis is error-prone and slow.
- **Goal:** Develop a hybrid deep-learning model for real-time detection.

Research Gap

Existing solutions struggle with high-dimensional, unlabeled data streams.

Literature Review

- **Context:** Modern distributed systems generate massive logs.
- **Pain Point:** Manual analysis is error-prone and slow.
- **Goal:** Develop a hybrid deep-learning model for real-time detection.

Research Gap

Existing solutions struggle with high-dimensional, unlabeled data streams.

Research Design

- **Context:** Modern distributed systems generate massive logs.
- **Pain Point:** Manual analysis is error-prone and slow.
- **Goal:** Develop a hybrid deep-learning model for real-time detection.

Research Gap

Existing solutions struggle with high-dimensional, unlabeled data streams.

Expected Contribution

- **Context:** Modern distributed systems generate massive logs.
- **Pain Point:** Manual analysis is error-prone and slow.
- **Goal:** Develop a hybrid deep-learning model for real-time detection.

Research Gap

Existing solutions struggle with high-dimensional, unlabeled data streams.

Proposed Mathematical Model

We assume the state transition follows a Markov process. The posterior probability is updated as follows:

$$p(\mathbf{x}_k|\mathbf{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Z}_{k-1})}{\int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Z}_{k-1})\, d\mathbf{x}_k} \tag{1}$$

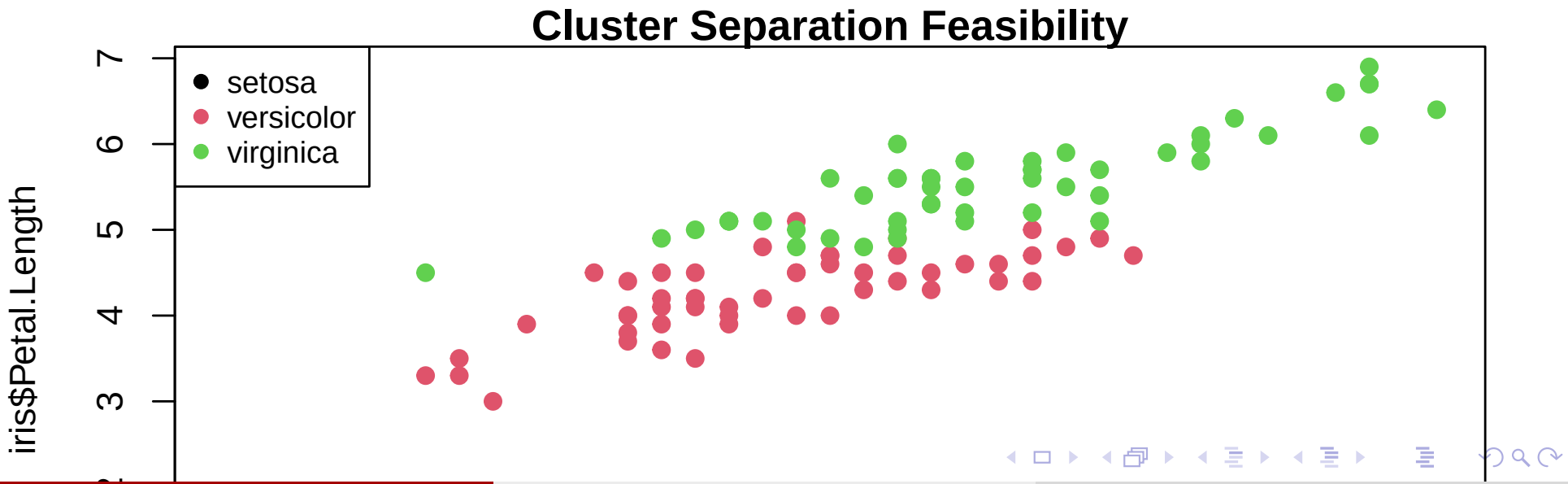
Where:

- $\mathbf{x}_k$ is the system state at time k
- $\mathbf{Z}_k$ is the observation history

Preliminary Data Analysis (R Integration)

Using R directly in slides ensures plots are always up-to-date with your data.

```r
# We can generate plots dynamically
par(mar=c(4,4,1,1))
plot(iris$Sepal.Length, iris$Petal.Length,
     col=iris$Species, pch=19,
     main="Cluster Separation Feasibility")
legend("topleft", legend=levels(iris$Species), col=1:3, pch=19
```



Project Timeline

Phase	Deadline
Literature Review	Month 1
Data Collection	Month 2
Implementation (Prototype)	Month 3-4
Evaluation & Tuning	Month 5
Writing Thesis	Month 6

Table: Estimated milestones for the 6-month period.

Next Steps

Thank you for your attention.

Questions?