The Best of Both Worlds? – Exploring the Use of Qualitative Analysis and Machine Learning Techniques on Academic Advising Data

T: +27 51 401 7421 | E: ctl@ufs.ac.za | **www.ufs.ac.za/ctl**

Thingo Mafaesa
Lucretia Bock

Inspiring excellence, transforming lives through quality, impact, and care.

Siyaphumelela
we succeed

VISION 130
Renew and Reimagine for 2034

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS

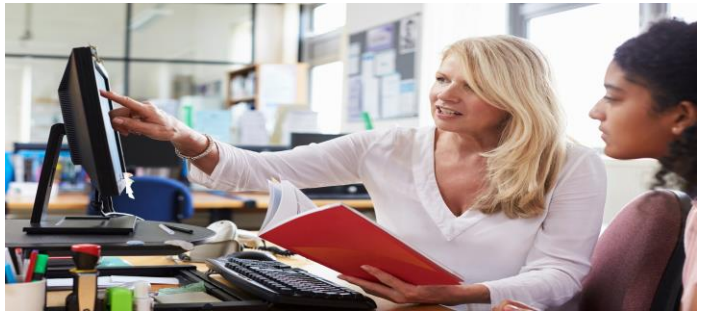
OUTLINE

- INTRODUCTION
- QUALITATIVE DATA ANALYSIS
- MACHINE LEARNING
- PROPOSED APPROACH
- CONCLUSION

Academic Advising

Academic advising is a **collaborative relationship** between a **student** and **an academic advisor**. The intent of this collaboration is to **assist the student** in the development **of meaningful educational goals** that are consistent with **personal interests, values and abilities**.

Academic advising increases students' **awareness of institutional support services** that are available at their disposal, as well as providing them with **information that is critical** in their educational journey. Academic advising also assists students, especially first years, in **transitioning** from basic education to the more complex and dynamic higher education environment.

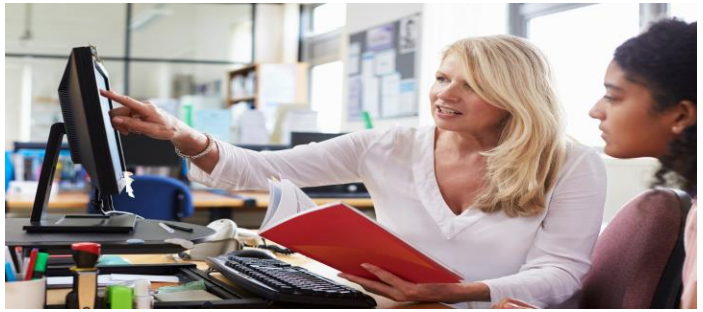


INTRODUCTION

The **Learner Case Management (LCM)** system is a tool UFS uses to capture qualitative data relating to academic advising sessions held between students and academic advisors.

There is a need for Qualitative Data Analysis, but why?

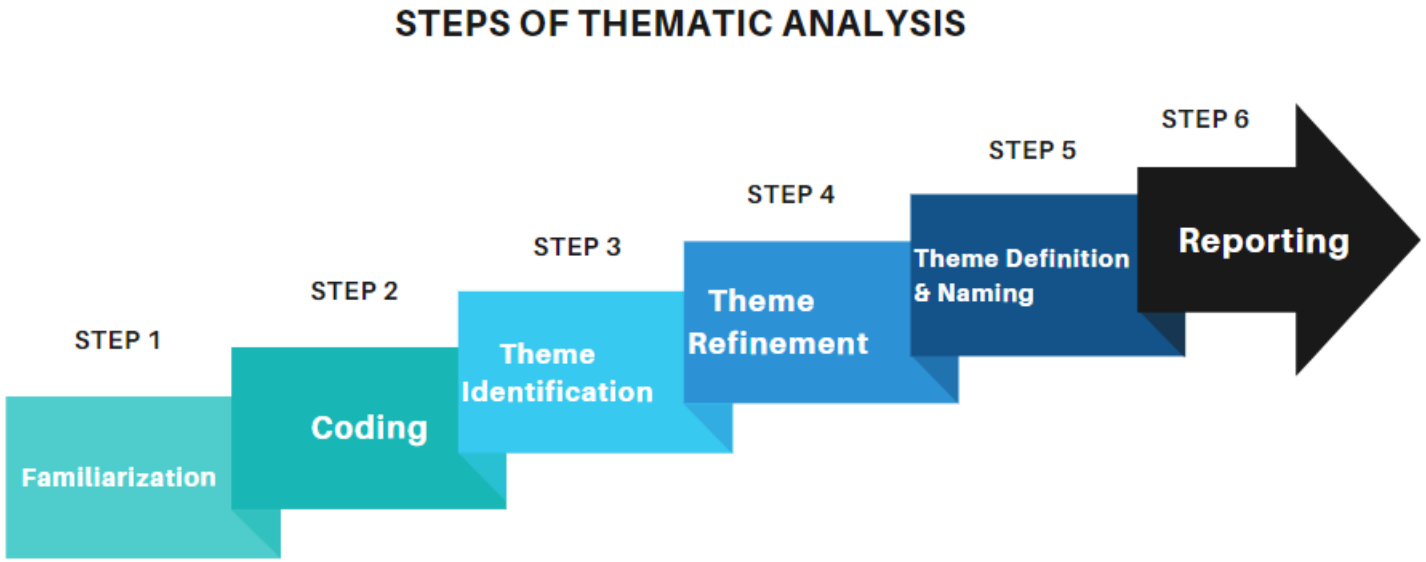
- To gain an in-depth understanding of the data
- To enhance data interpretation and reporting
- To improve interventions
- To support evidence-based decision making



THEMATIC ANALYSIS

What is Thematic Analysis?

Thematic analysis is a method for analysing qualitative data by thoroughly **reviewing the data set** and **identifying** recurring patterns or **themes**. This method requires the researcher to actively reflect, placing their insights and experiences at the forefront of interpreting the data. (Braun & Clarke, 2006)



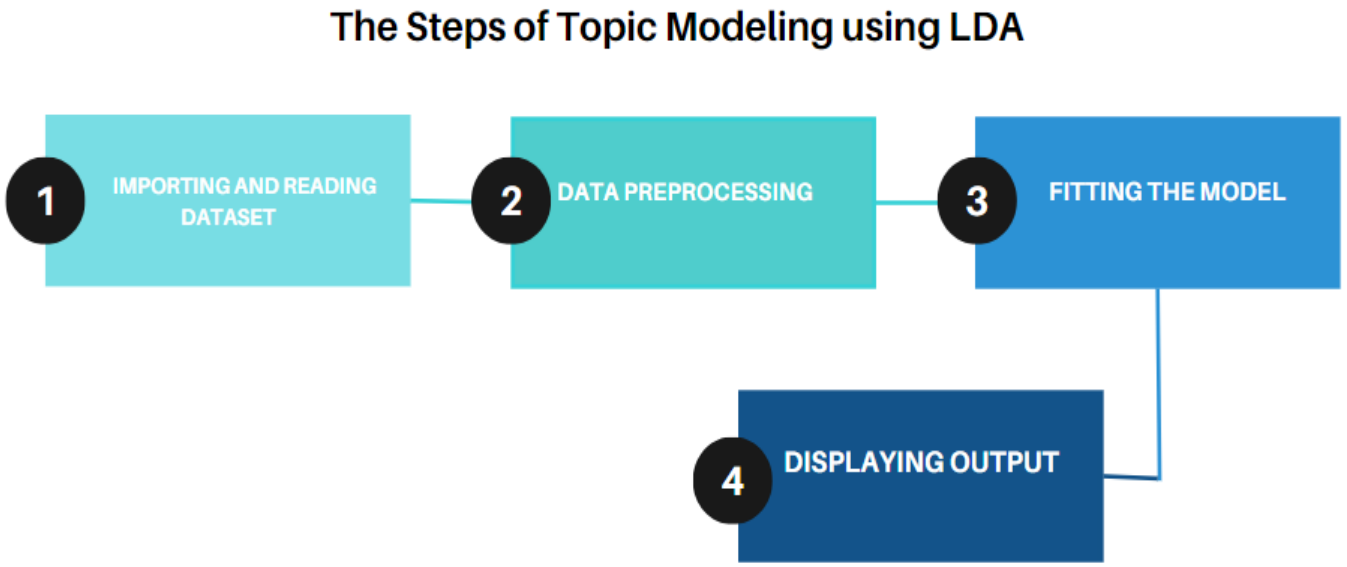
CHALLENGES

- **Time-Intensive**: Thematic analysis can be very time-consuming, requiring repeated reading and coding of data, which can be resource-intensive.
- **Subjectivity and Bias**: The researcher's interpretation is central to identifying themes, which can introduce personal bias and affect the objectivity of the analysis.
- **Data Saturation**: Determining when data saturation is reached— when no new themes are emerging—can be subjective and uncertain.
- **Depth vs. Breadth**: Balancing the depth and breadth of analysis can be challenging; focusing too much on specific themes may overlook broader patterns, and vice versa.

Topic Modelling - LDA

LATENT DIRICHLET ALLOCATION (LDA)

Latent Dirichlet Allocation (LDA) is a **machine-learning natural language processing** method used for **topic modelling.** For a given collection of documents or data sets, it identifies the **main topics** and **their distributions** by applying a **probabilistic model** that assigns words to these topics.



EXAMPLE

PYTHON

```python
# packages to store and manipulate data
import pandas as pd
import numpy as np

# plotting packages
import matplotlib.pyplot as plt
import seaborn as sns

# model building package
import sklearn

# package to clean text
import re
import nltk
from nltk.tokenize import RegexpTokenizer
from nltk.corpus import stopwords
```

```python
from sklearn.feature_extraction.text import CountVectorizer

# the vectorizer object will be used to transform text to vector form
vectorizer = CountVectorizer(max_df=0.9, min_df=25, token_pattern='\w+|\$[\d\.]+|\S+')

# apply transformation
tf = vectorizer.fit_transform(df['clean_narrative']).toarray()

# tf_feature_names tells us what word each column in the matric represents
tf_feature_names = vectorizer.get_feature_names_out()

from sklearn.decomposition import LatentDirichletAllocation
from sklearn.decomposition import NMF

number_of_topics = 10

model = LatentDirichletAllocation(n_components=number_of_topics, random_state=0)
#model = NMF(n_components=number_of_topics, random_state=0)

model.fit(tf)
```

SAMPLE OUTPUT

Topic 1		Topic 2		Topic 3	
term	weight	term	weight	term	weight
game	0.014	space	0.021	drive	0.021
team	0.011	nasa	0.006	card	0.015
hockey	0.009	earth	0.006	system	0.013
play	0.008	henry	0.005	scsi	0.012
games	0.007	launch	0.004	hard	0.011

DATA

Advice Narratives: 2022 - 2023

750 Narratives

Themes and Codes

Thematic Analysis

Academic Challenges
- Heavy workload
- Study methods
- Time management
- Tutorial assistance required

Administration

Career
- Career path
- Limited Opportunities

Environment
- Accommodation
- Adapting to new environment
- Transfer to another campus

Finances
- Lack of funding
- NSFAS Propensity Letter

Personal Challenges

Wellbeing
- Health issues affecting studies
- Mental Health

RESULTS – LDA (Topic Modelling)

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
assist	student	would	also	student	time	major	student	want	student
studi	refer	modul	discuss	modul	manag	chang	want	student	semest
student	campu	like	consid	regist	struggl	advisor	degre	degre	modul
plan	faculti	year	possibl	want	student	psycholog	educ	know	support
need	due	exam	b	faculti	studi	explain	programm	chang	first
strategi	challeng	write	psycholog	year	schedul	want	career	transfer	regist
academ	well	understand	appli	rule	help	work	scienc	b	studi
modul	person	fail	session	fail	assist	interest	option	modul	schedul
set	regard	student	look	advic	also	option	chang	could	advis
appoint	requir	semest	complet	given	need	student	would	issu	session
Topics									
Study strategies	-	Exam preparations	Mental health	Registration	Time management	Program/major change	Change degree program	Change degree program	Tutorial Support needed

Inspiring excellence, transforming lives through quality, impact, and care.

Siyaphumelela
we succeed

VISION 130
Renew and Reimagine for 2034

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS

RESULTS – LDA (Topic Modelling) cont. - Improvements

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
academ	faculti	studi	discuss	degre	issu	want	semest	manag	modul
advic	know	time	nsfa	chang	receiv	psycholog	cours	time	semest
major	want	manag	letter	want	session	interest	studi	appoint	fail
modul	modul	assist	continu	programm	need	educ	first	well	year
registr	indic	need	option	appli	support	advisor	find	challeng	want
chang	rule	strategi	modul	current	regist	career	also	requir	regist
given	book	follow	given	scienc	inform	degre	struggl	refer	second
request	year	struggl	rule	transfer	career	chang	year	academ	first
assist	complet	plan	academ	possibl	letter	explain	possibl	advisor	exam
regard	refer	schedul	inform	account	find	major	need	assist	write
Topics									
Registration	Information	Time management/study strategies	Funding	Degree change	Registration/Career change	Degree change	Information	Time management	Registration/Exam prep

ADVANTAGES - LDA

Time efficiency

Reduction of bias

Ability to process large amounts of data with no pre-defined vocabulary

CHALLENGES - LDA

- **Sensitive to the number of topics chosen** – Choosing a different number of topics may lead to different topics
- **Repetitive topics** – Sometimes leads to similar topics
- **Works better with large datasets**
- **Contextual meaning is still required** – Subject experts are needed.
- **Sometimes creates topics that do not make sense**
- **Requires analytical skillset** – not readily available

COMPARING TA THEMES TO LDA TOPICS

Themes and Codes

Academic Challenges
- Heavy workload

- Study methods
- Time management
- Tutorial assistance required

Administration

Career
- Career path

Personal Challenges

Environment
- Accommodation
- Adapting to new environment
- Transfer to another campus

Finances
- Lack of funding

- NSFAS Propensity Letter

Wellbeing
- Health issues affecting studies
- Mental Health

LDA Topics

Study Strategies
Time Management
Exam preparations
Tutorial Support

Registration
Information

Program/major change
Change degree program

Mental Health

Funding

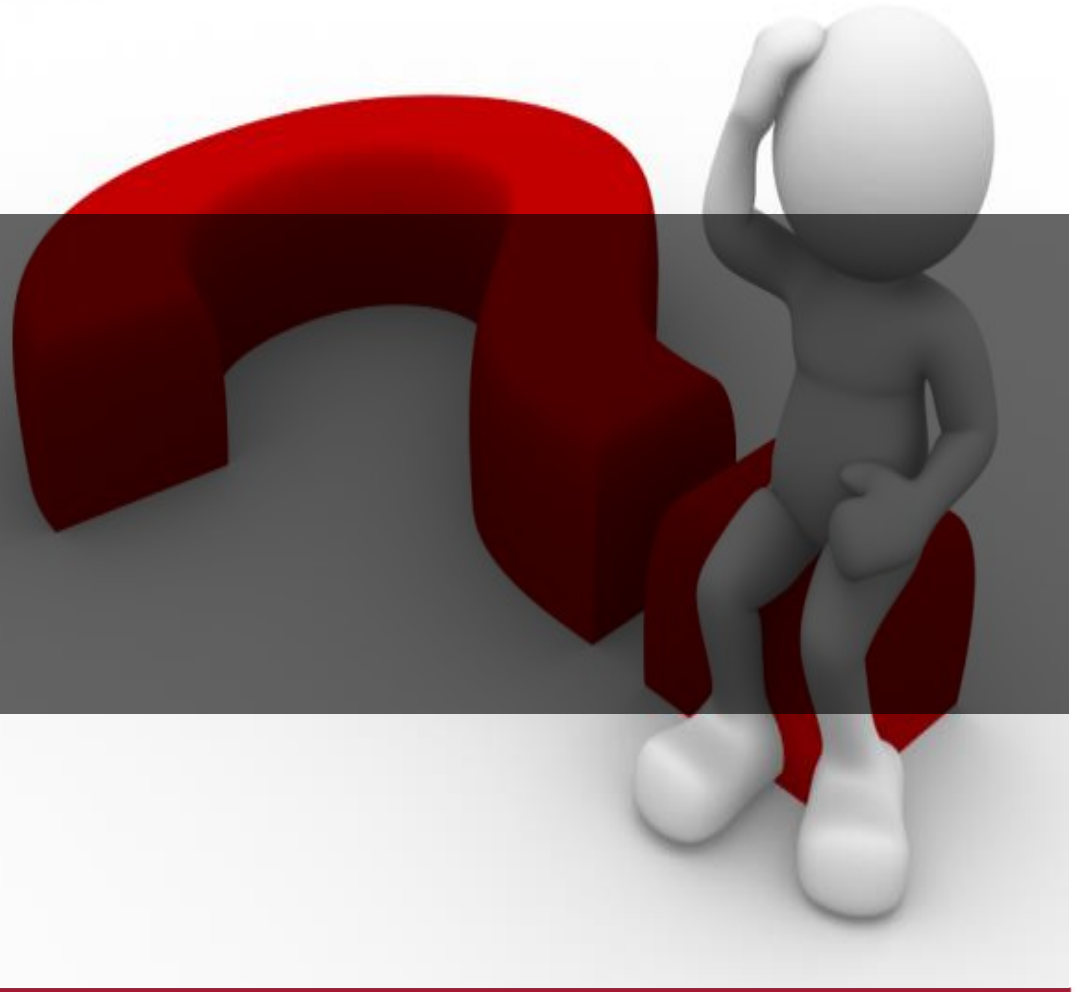
Data Analysts, Research Analysts & Those who want results!

A Combined Approach ~~(Mixed method)~~

- ✓ The **LDA process** *replaces Step 1 and Step 2* of **thematic analysis**
- ✓ It is **time efficient**
- ✓ A **balance** between **objectivity** and **subjectivity** in the results
- ✓ **Improved data interpretation**: Qualitative analysis adds context to quantitative results, making it easier to interpret numerical data within real-world settings and conditions.

Whether you will admit it to yourself or not, ***we are all creatures of comfort***. We want to get more done, quicker, and easier, while still maintaining quality!

Thank you



T: +27 51 401 7421 | E: ctl@ufs.ac.za | **www.ufs.ac.za/ctl**

Inspiring excellence, transforming lives through quality, impact, and care.

Siyaphumelela
we succeed

VISION 130
Renew and Reimagine for 2034

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS