

Unveiling the Voices of Tanzania Farmers: A Text Analysis of Open-Ended Responses from the National Sample Census of Agriculture

Gabriel Kulomba Simbila

Email: gkulomba@gmail.com

Tanzania

Abstract Confirmation Number: 3805

Abstract

Tanzania's agricultural sector is pivotal to its economy and food security, yet farmers continue to face multifaceted challenges that hinder productivity and sustainable development. While quantitative data from the National Sample Census of Agriculture (NSCA), conducted every five years by the National Bureau of Statistics (NBS), provide valuable insights, the open-ended responses remain underutilized for understanding farmers' lived experiences.

This research paper systematically analyzes qualitative responses from the 2019/2020 NSCA using natural language processing (NLP), thematic analysis, topic modeling, and sentiment analysis. The findings reveal key challenges related to access to inputs, market constraints, climate variability, institutional support gaps, and infrastructure deficiencies.

By amplifying farmers' voices, this research provides policy-relevant evidence to inform targeted, inclusive, and geographically responsive agricultural interventions in Tanzania.

Keywords: *Farmer perspectives; open-ended survey responses; text analysis;*

1. Introduction and Background

Agriculture is the backbone of Tanzania's economy, employing about 65% of the labor force and playing a vital role in food security and poverty reduction. However, farmers face persistent challenges such as limited access to inputs, markets, extension services, finance, and increasing climate variability. Although national surveys provide extensive quantitative data, they often overlook farmers' lived experiences and qualitative realities.

2. Problem Statement

Although the NSCA includes open-ended questions, these qualitative responses are rarely analyzed systematically. Most assessments focus on quantitative indicators such as production levels and input utilization, creating an evidence gap regarding farmers' narratives and lived experiences.

This limits the ability of agricultural policies and programmes to fully reflect farmers' priorities, particularly concerning market access, extension services, climate risks, and institutional support. Advances in computational text analysis offer scalable tools to integrate qualitative insights with quantitative data, yet these methods remain underutilized in Tanzania's agricultural policy analysis.

Footnote: *The authors confirm that this manuscript contains original work and does not violate any copyright or intellectual property rights*

3. Research Questions

This study addresses the following questions:

1. What are the predominant challenges reported by Tanzanian farmers in NSCA open-ended responses?
2. What additional insights do these narratives provide beyond closed-ended survey indicators?
3. How do agricultural constraints vary across regions and agro-ecological zones?
4. How do gender and youth perspectives differ in reported constraints?
5. How do population pressure and land scarcity influence farming practices and vulnerability?

4. Literature Review

The literature emphasizes agriculture's role in rural transformation and climate resilience in sub-Saharan Africa (Pretty et al., 2011; Below et al., 2012). Studies highlight the importance of socio-economic characteristics, institutional access, and gender dynamics in shaping productivity (Mkenda et al., 2019; Sanga et al., 2020).

Methodologically, mixed-methods approaches are increasingly used to integrate quantitative and qualitative insights (Creswell & Plano Clark, 2018). Advances in computational text analysis—such as thematic coding, topic modeling, and automated content analysis—enable scalable analysis of large datasets (Blei et al., 2003; Braun & Clarke, 2006; Grimmer & Stewart, 2013).

5. Data and Methodology

5.1 Data Source

The study uses open-ended responses from the 2019/2020 NSCA, covering approximately 95,000 agricultural households across all regions of Mainland Tanzania. Responses capture farmers' perceptions of constraints, opportunities, and priority interventions. Text data were anonymized, cleaned, and standardized prior to analysis.

5.2 Analytical Approach

A mixed computational and thematic framework was applied:

Text Pre-processing

- Tokenization
- Lower-casing
- Stop-word removal
- Stemming/Lemmatization
- Spelling normalization

Descriptive Text Analysis

- Word frequency analysis
- N-gram identification

Word Cloud Visualization

- Frequency-based visual representation

Sentiment Analysis

- Classification into positive, negative, and neutral categories

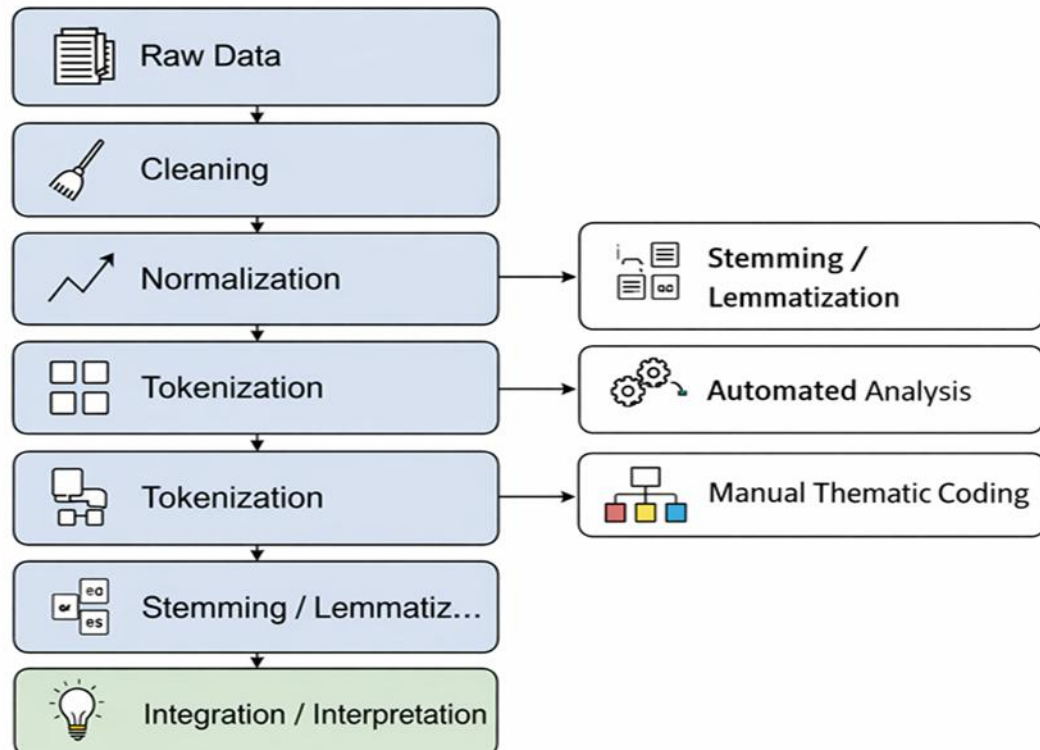
Topic Modeling (LDA)

- Extraction of latent thematic clusters

Disaggregation

- Gender
- Youth (15–35 years)
- Region
- Agro-ecological zone

Figure 1. Workflow of Text Analysis



Raw Data → Cleaning → Normalization → Tokenization → Stemming/Lemmatization → Automated Analysis → Manual Coding → Interpretation

6. Results

6.1 Geographic Coverage

Figure 2. Geographic Coverage of the 2019/2020 NSCA



The map illustrates nationwide coverage across Mainland Tanzania, ensuring regional representation and agro-ecological diversity.

6.2 National-Level Challenges

Word Cloud (Figure 3)

Figure 2. Word Cloud of Most Frequent Terms in Farmers' Responses (All Ages)



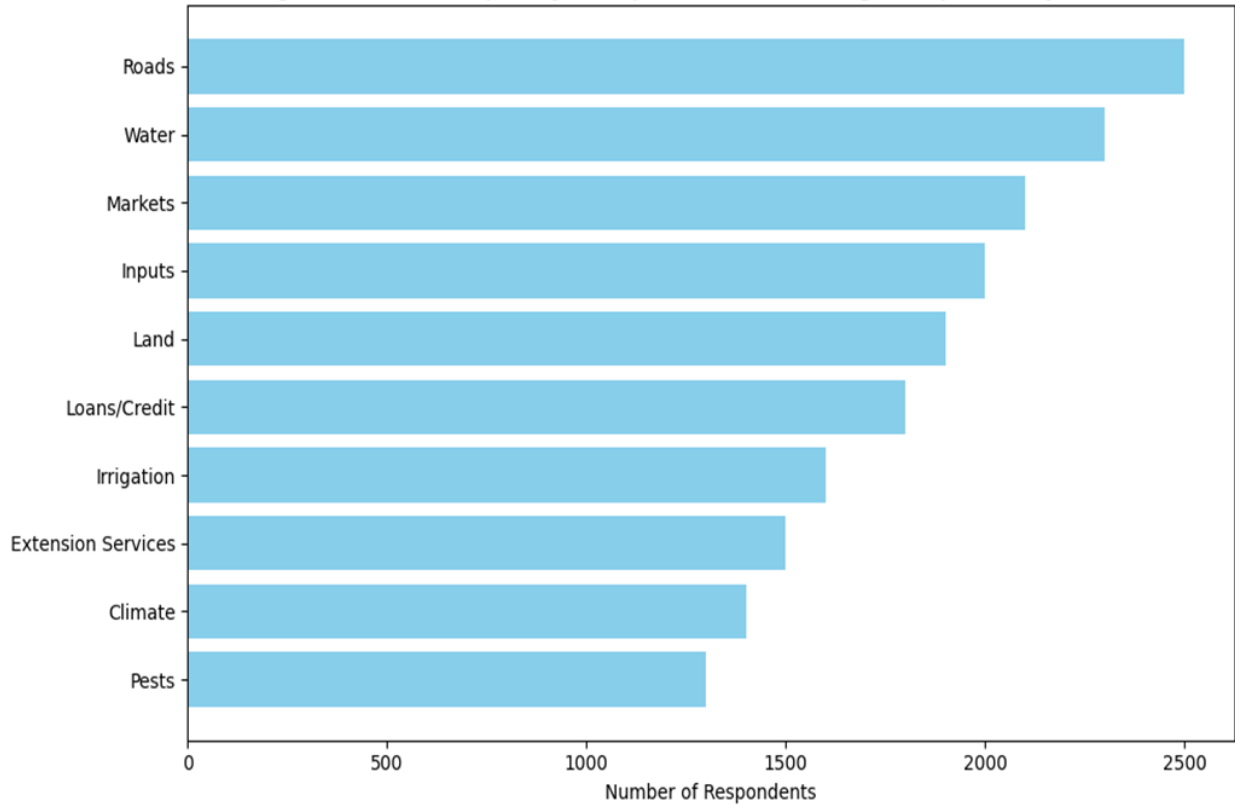
The national word cloud highlights dominant concerns including:

- Roads
- Water access
- Markets
- Inputs
- Land
- Loans
- Irrigation
- Extension services
- Climate
- Pests
- Fertilizer

Word size reflects frequency, providing an intuitive overview of national-level priorities.

Term Frequency Bar Chart (Figure 4)

Figure 3. Term Frequency of Top National Challenges Reported by Farmers



Top national challenges include:

- Roads
- Water
- Markets
- Inputs
- Land

This complements the word cloud with quantitative clarity.

6.3 Sentiment Analysis (Figure 5)

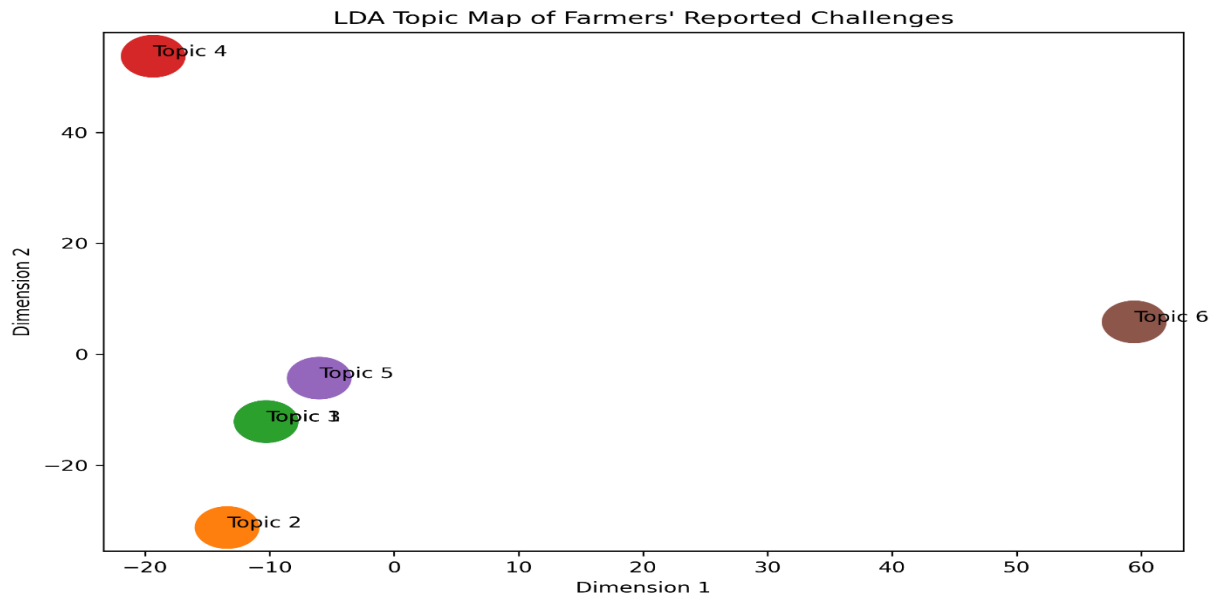


Negative sentiment dominates, particularly regarding:

- Climate shocks
- High input costs
- Poor infrastructure
- Limited market access

Neutral sentiments relate to descriptive statements; positive sentiments are limited and often linked to extension services or cooperative participation.

6.4 LDA Topic Map Showing Clusters of Themes in Farmers' Responses

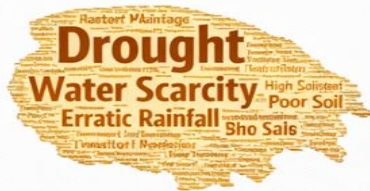


The LDA topic map of NSCA open-ended responses reveals key clusters of farmers' challenges across Mainland Tanzania: input and production constraints, climate and environmental risks, market and economic barriers, and infrastructure and services. Circle size shows prevalence, and proximity indicates relatedness, highlighting how these issues are interconnected. This visualization complements quantitative analyses by providing an intuitive overview of multidimensional agricultural constraints.

6.5 Agro-Ecological Zone Analysis (Figure 7)

Word Cloud Maps of Agricultural Challenges by Ecological Zones - Mainland Tanzania

Semi-Arid Central Zone



Northern Highland and Volcanic Zone



Lake and Western Zones



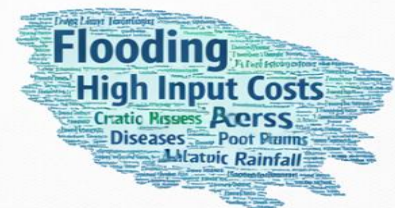
Southern Highland Zone











Southern Highland Zone



Coastal and Eastern Zone



6.7 Gender-Specific Constraints

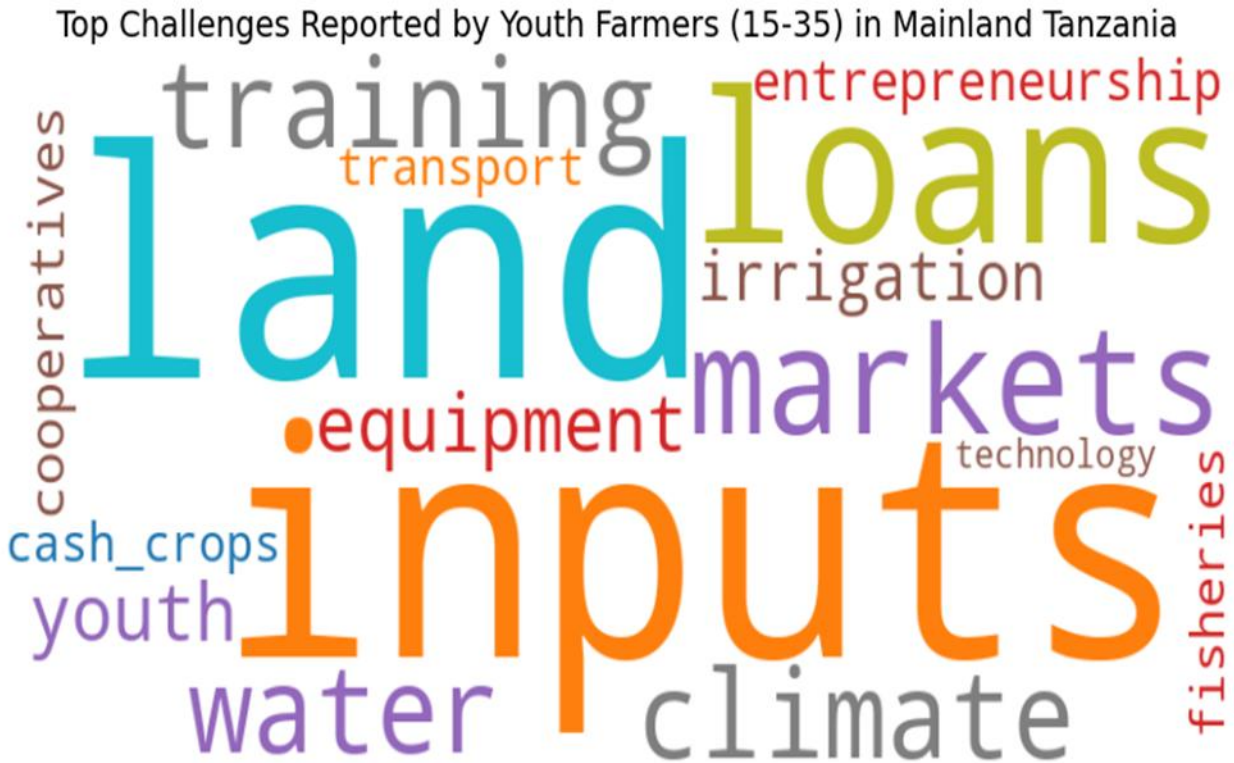
Gendered Constraints and Policy Implications from Farmers' Responses			
Thematic Area	Issues for Women	Issues for Men	Policy Recommendations
 Access to Land	Limited ownership; Insecure inheritance	Greater control; Land preparation	Strengthen women's land rights; Joint land titling
 Access to Inputs	Low income; Limited subsidies	Higher input use; Cost constraints	Target women in input programs; Support agro-dealers
 Access to Finance	Lack of collateral; Depend on savings groups	More formal credit access	Expand women's finance options, Introduce credit guarantees
 Labour & Time Burden	Heavy workload; Time poverty	Seasonal labour gaps	Promote labour-saving tech; Rural childcare support
 Extension & Markets	Limited mobility; Farm gate sales	Better market access; Bargaining power	Support women's cooperatives; Improve transport
 Technology Adoption	Low tech adoption; Access barriers	Higher tech adoption	Provide training & grants for women
 Climate Challenges	Drought impact; Food security	Yield loss concerns; Livestock risks	Promote climate-smart practices; Water harvesting
 Decision-Making	Limited income control	Control of farm revenues	Encourage joint decision-making; Women's producer groups

Women report:

- Heavy domestic workloads
- Limited access to land and finance
- Barriers to market participation

These constraints reduce productivity and technology adoption.

6.8 Youth Perspectives (Figure 9)



Youth (15–35) face:

Limited land access

Credit constraints

Training gaps

Market barriers

However, youth demonstrate interest in technology, entrepreneurship, and cooperatives.

6.9 Population Pressure and Land Scarcity (Figure 10)



Land fragmentation and scarcity—especially in northern Tanzania—amplify vulnerability and reduce resilience to climate and market shocks.

7. Discussion and Policy Implications

Findings reveal widespread dissatisfaction with infrastructure, irrigation, inputs, and extension services.

Regional and demographic disparities underscore the need for:

Targeted regional interventions

Gender-responsive agricultural policies

Youth-focused land and finance programs

Climate-smart agriculture

Improved rural infrastructure

Integrating qualitative census data strengthens evidence-based planning. Conclusion

Text analysis of NSCA open-ended responses reveals persistent rural challenges driven by limited market access, unstable crop prices, inadequate extension services, and infrastructural deficits. Farmers' narratives highlight regional disparities as well as gender- and youth-specific constraints. This study demonstrates the power of text analysis in amplifying farmers' voices and providing policy-relevant evidence for designing inclusive, context-sensitive agricultural policies in Tanzania.

8. Conclusion

Text analysis of NSCA open-ended responses reveals persistent rural constraints shaped by market barriers, climate variability, infrastructure deficits, and institutional gaps. Farmers' narratives highlight regional diversity and gender- and youth-specific vulnerabilities.

This research demonstrates the power of computational text analysis in amplifying farmers' voices and generating policy-relevant evidence for inclusive and context-sensitive agricultural development in Tanzania.

References (Vancouver Style)

1. Pretty J, Toulmin C, Williams S. Sustainable intensification in African agriculture. *Int J Agric Sustain*. 2011;9(1):5–24.
2. Below R, Mutabazi K, Kirschke D, Franke C, Sieber S, Tscherning K. Can farmers' adaptation to climate change be explained by socio-economic household-level variables? *Glob Environ Change*. 2012;22(1):223–235.
3. Mkenda BK, et al. Policy processes and agricultural transformation in Tanzania. *Afr Dev Rev*. 2019;31(S1):50–64.
4. Sanga U, Mlozi M, Haug R. Gendered access to extension services in Tanzania. *J Agric Educ Ext*. 2020;26(3):247–264.
5. Food and Agriculture Organization of the United Nations. *Agricultural censuses and surveys: Guidelines and tools*. Rome: FAO; 2018.
6. National Bureau of Statistics Tanzania. *National Sample Census of Agriculture 2014/15*. Dar es Salaam: NBS; 2016.
7. United Republic of Tanzania. *Agriculture Sector Development Programme Phase II (ASDP II)*. Dodoma: Government of Tanzania; 2019.

8. Creswell JW, Plano Clark VL. Designing and conducting mixed methods research. 3rd ed. Thousand Oaks (CA): Sage Publications; 2018.
9. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs. *Ann Fam Med*. 2013;11(2):213–222.
10. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77–101.
11. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *J Mach Learn Res*. 2003; 3:993–1022.
12. Grimmer J, Stewart BM. Text as data: The promise and pitfalls of automatic content analysis methods. *Polit Anal*. 2013;21(3):267–297.