ATHENA

Leveraging Artificial Intelligence and Big Data for **IFAD 2.0** – Phase 2

IFAD
Investing in rural people

Leveraging Artificial Intelligence and Big Data for **IFAD 2.0** – Phase 2

The "Athena" project – Final Report

February 2021

IFAD
Investing in rural people

Contents

List of tables and figures

Acknowledgements

The second phase of the Innovation Challenge project "Leveraging Artificial Intelligence and Big Data for IFAD 2.0", also termed the "Athena" project, was led by Alessandra Garbero, Senior Economist in the Research and Impact Assessment Division (RIA).

The following consultants worked towards the success of this project (in alphabetical order): Bia Carneiro, Arthur Charpentier, Giovanni Cerulli, Camila Comunello, Daniele di Simone, Ewen Gallic, Maryia Ivanina, Aliaksei Ivanin, Jaron Porciello, Giuliano Resce, and Grayson Sakos.

The work benefited from technical inputs from the Sustainable Production, Markets and Institutions Division (PMI), notably from the Director Thouraya Triki, and PMI staff and consultants: Richard Abila, Jean-Philippe Audinet, Mawira Chitima, Robert Delve, Michael Hamp, Enrique Hennings, Mylene Kherallah, Rikke Grand Olivera, Antonio Rota, Kathy Zissimopoulos, Simone Sala and Valeria Silvestri as well as ECG staff and consultants (Beatrice Gerli). Consultations were also held with Programme Management Department (PMD) Front Office and OPR staff. Inputs were also received from the Knowledge Management Group in the Strategy and Knowledge Department (SKD). Support was also provided by the Information Communication Technology (ICT) and Administrative Services Divisions (ADM).

Additional people contributed to this project, notably Tuuli Sauren, Anna Gloria Bille and John Laird. A special acknowledgement therefore goes to all those who worked on this project as well as to those who made this possible, notably the Change Delivery and Innovation Unit (CDI), Edward Gallagher and Gladys Herminia Morales. Lastly, we are grateful to the RIA Director, Sara Savastano, and Associate Vice-President, Strategy and Knowledge Department, Meike Van Ginneken, for supporting this project.

Acronyms and abbreviations

AI	artificial intelligence
COSOP	country strategic opportunities programme
GRIPS	Grant and Investment Projects System
IA	Impact assessment
ICT4D	information and communications technology for development
ML	machine learning
MTR	midterm review report
ORMS	Operational Results Management System
PCR	project completion report
PDR	project design report
RDR	Rural Development Report
SDG	Sustainable Development Goal

ATHENA

Executive Summary

This report is the output of the second phase of the 2019 IFAD Innovation Challenge project "Leveraging Artificial Intelligence and Big Data for IFAD 2.0", led by Alessandra Garbero, Senior Economist within the Research and Impact Assessment Division (RIA). The project was named "Athena" as she is the goddess of wisdom in ancient Greek mythology. Hence the analogy with artificial intelligence, a wide-ranging branch of computer science that has the power to harness knowledge within organizations. Its second phase was implemented between July and November 2020.

In this phase the project team: (i) refined the algorithms and cross-validated the models and results with different sets of data; and (ii) harmonized the models with existing IFAD frameworks to establish an artificial intelligence (AI) "tool box" that contains:

- **AI-based Intervention Dashboard:** a searchable dashboard that classifies IFAD's investment portfolio for several project features, such as interventions, outcomes, animal and plant products, among others;
- **Lessons Learned Web App**: an application to search for topics from "lessons learned" as reported in project completion reports (PCRs);
- **Trend analyses of strategic themes**: historical evidence of activities related to strategic topics (such as Sustainable Development goals [SDGs]), food systems and information and communications technology for development [ICT4D];
- **Project performance prediction model**: a framework for ex-ante prediction of project performance based on a set of features;
- **Project impact prediction model**: framework to predict the probability of a positive impact of IFAD-supported interventions;
- **Project targeting optimization model**: framework to identify beneficiary features to maximize project impact; and
- **Covid-19 impact prediction model**: framework to predict impact of the pandemic in IFAD's beneficiary countries.

The AI-based intervention dashboard was tested and validated by several key stakeholders within IFAD, having received positive feedback with regards to its applicability for knowledge generation. The other tools are to be tested in phase 3.

In the second phase, machine learning (ML) was trialled to systematise and integrate different data types and sources in order to accelerate knowledge generation. This has several implications from a policy perspective and can provide added value in:

(1) **Aiding and simplifying IFAD reporting,** for instance, in the Report on IFAD's Development Effectiveness (RIDE) – but also for more complex and data hungry thematic reporting (such as mainstreaming themes, food systems components and extent of reporting against SDG targets and goals in IFAD projects).

(2) **Enabling ex-ante data driven design**: **informing the design of new operations particularly country strategic opportunities programmes (COSOPs), and project design documents** about cost-effectiveness of interventions proposed as well as impact and performance potential.

(3) **Inform ex-ante about targeting strategies** in design documents by proposing a menu of options for assigning projects interventions in order to have impactful operations.

(4) **Enhancing development effectiveness** by contributing data driven and evidence based analytical solutions at each step of the project cycle.

(5) **Supporting the ICT4D development strategy** which foresees AI, ML and big data as key pillars.

(6) **Contribute to the knowledge management action plan**, by aiding the dissemination of knowledge to users in a user-friendly way via the development of dashboards, apps and automated briefs.

Overview

In the first phase of the 2019 IFAD Innovation Challenge, the “Athena” project team sought to unlock the potential of artificial intelligence (AI) to accelerate knowledge generation and strengthen data-driven decision-making in IFAD. An IFAD team led by staff from the Research and Impact Assessment Division (RIA) composed of economists, data scientists and social scientists explored machine learning (ML) techniques to extract insights from IFAD investments. We tested a number of innovations and proposed an integrated, machine-driven approach to analyse project documentation and predict impact.

In the second phase of the project (the subject of this report), we refined the algorithms and cross-validated the models and results with different sets of data. We harmonized the models with existing IFAD frameworks to establish an AI “tool box” that contains:

- **AI-Based Intervention Dashboard:** a searchable dashboard that classifies IFAD’s investment portfolio for several project features, such as interventions, outcomes, animal and plant products, among others;
- **Lessons Learned Web App**: an application to search for topics from Lessons Learned reported in project completion reports (PCRs).
- **Trends analysis of strategic themes**: historical evidence of activities related to strategic topics (such as Sustainable Development Goals [SDGs], food systems, ICT4D).
- **Project performance prediction model**: framework for ex-ante prediction of project performance based on a set of features.
- **Project impact prediction model**: framework to predict the probability of positive impact of IFAD-supported policies.
- **Project targeting optimization model**: framework to identify beneficiary features to maximize project impact.
- **Covid-19 impact prediction model**: framework to predict impact of the pandemic in IFAD beneficiary countries.

With these various applications, the project aimed to accomplish **two** key objectives:

(i) **First, to understand and systematize IFAD’s investment portfolio since 1981,** using a variety of datasets, both quantitative and qualitative (specifically text from project reports), in order to determine the distribution of themes, interventions, development outcomes and lessons learned, as well as the extent of reporting towards strategic topics such as the SDGs and food systems components.

(ii) **Secondly, to set up the infrastructure for predictive analytics,** through the development of algorithms that support the project cycle by, for instance, highlighting features of successful portfolio performance or determining the beneficiary and project level features that drive positive impact.

This global overview of IFAD’s investments and an understanding of what drives project success aims to enable the main goal of this innovation project, which is **to enhance and accelerate knowledge management**. Figure 1 presents the overall conceptual framework for the project’s approach, with the various techniques and applications. This report summarises the main results for each application and use case.

Machine learning models for data-driven decision making

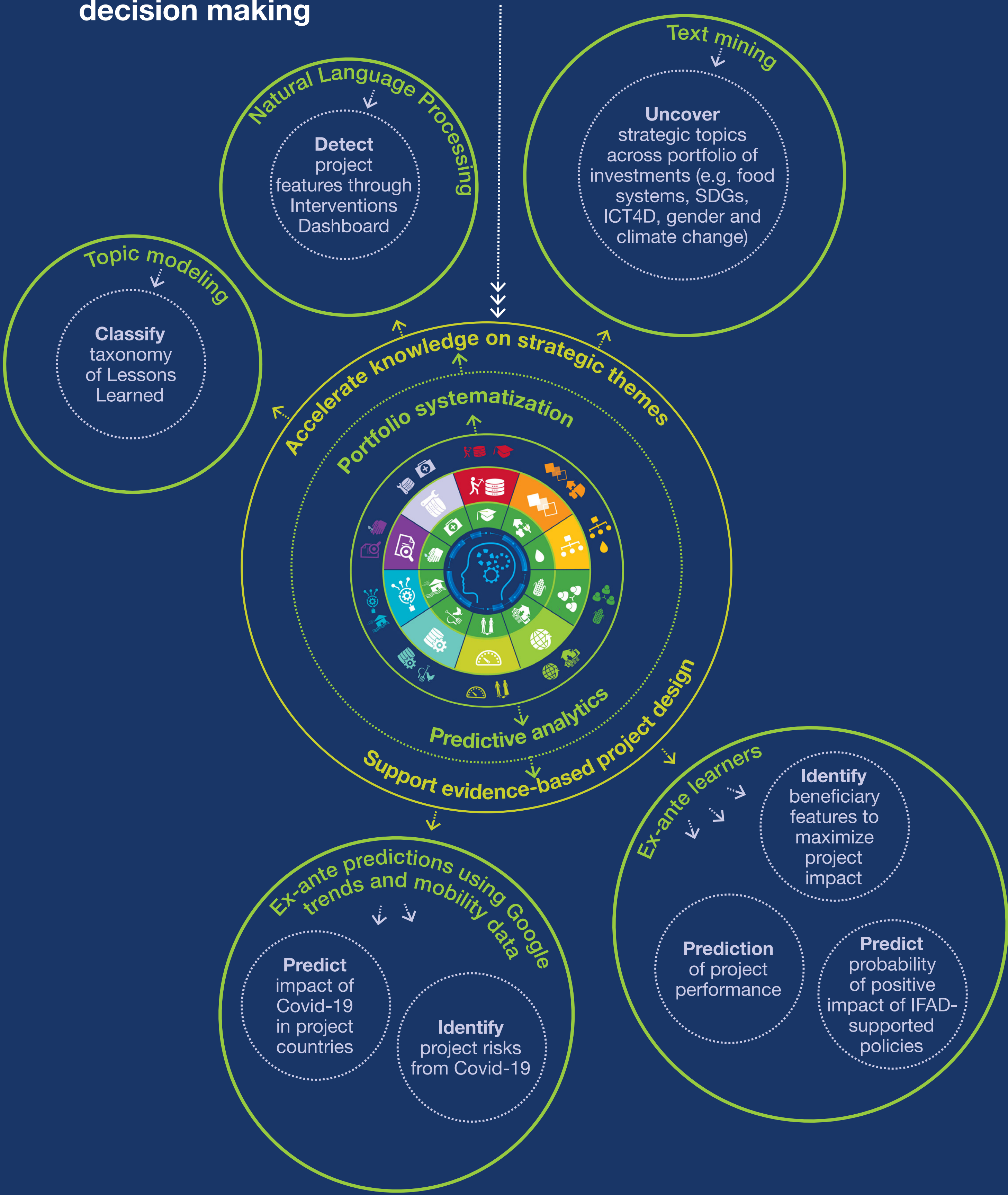


Figure 1 Conceptual framework.

Main results

Portfolio systematization

AI-Based Intervention Dashboard:[1] using machine learning to identify interventions, outcomes and other features in project documents

In phase 1, an advanced ML model, named Persephone[2] (designed specifically for agriculture) was adapted and applied to examine project documentation (mostly project design reports [PDRs]) for a total of 866 projects from 1980 to 2020, representing the totality of IFAD project investment portfolio since its foundation. In this phase we aimed at improving the reliability of the model by triangulating with other datasets and harmonizing with the new Grant and Investment Projects System (GRIPS) categorization framework. The final output was a public dashboard, which was tested by domain experts and other users and was commended for its utility.

Identifying and visualizing the distribution of granular interventions is the key focus of the dashboard. While IFAD classifies projects activities into broader categories (as described in the IFAD new categorization framework document[3]) granular interventions can be considered as the very specific activities that fall under the broader category and subcategories brackets. For instance, agricultural extension services, access roads, domestic water supplies, soil conservation, water users' associations, territorial planning, on-farm enterprises, education services, marketing support, employment generation, electricity, information access, pasture development, credit delivery, ground water resources, well usage and rural markets are possible examples. The list of intervention types was extracted from Persephone's list of 30,000 examples, which were then detected within IFAD project documentation using the ML algorithms. 5,000 unique interventions were identified. Through natural language processing (NLP) algorithms, the model classified not only these granular interventions, but also associated features such as: expected outcomes[4] (as stated in PDRs), presence of mainstreaming themes, animal and plant products,[5] among others. Each intervention was also mapped to the latest iteration of IFAD's GRIPS categories through a customization of the machine-learning model.

Table 1 below shows: the overall top 10 interventions found in project documents; the number of projects where the intervention is found (the "project count); and in how many sentences. The most frequent interventions were related to training, loans and financial services.

Table 1 Top 10 interventions identified in IFAD projects, by unique project count (n = 866 projects)

Intervention	Unique count of project ID	Number of sentences
training	615	18,610
loan	599	12,321
financial service	573	5,550
irrigation	570	9,218
infrastructure	560	9,140
investment	555	8,026
consultation	555	2,325
marketing	526	6,124
technical assistance	522	4,263
Cropping	514	5,060

1 Dashboard is available at this address: https://webapps.ifad.org/ria/dashboard Latest version of Dashboard User Guide (Nov 2020): https://www.dropbox.com/s/einch4ww9ydm740/Dashboard%20-%20user%20guide%20v5.docx?dl=0

2 Porciello, J., Ivanina, M., Islam, M. *et al.* (2020) Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. *Nat Mach Intell* 2, 559–565.

3 IFAD (2020). A new categorization framework for IFAD-supported project interventions. Operations Manual.

4 As impact assessments are not part of this dataset, but outcomes are still described in the documentation throughout the project cycle, query results should be treated as "expected outcomes" for the projects.

5 In accordance with FAO's multi-language AGROVOC thesaurus: http://aims.fao.org/vest-registry/vocabularies/agrovoc

Figure 2 shows the number of GRIPS categories found per project. It is important to note that these figures differ from the categories attributed to project components in IFAD's Operational Results Management System (ORMS) because the model classifies each granular intervention individually, according to category-specific taxonomies. The model has a 60% accuracy for classification on the 62 GRIPS subcomponent types and 70% accuracy for classification on 14 GRIPS category types. The three most frequent categories identified by the model, which are equally present in 853 projects are: crops, policy development and engagement, and rural business development. The category banking and financial services comes a close second, having been detected in 852 projects.

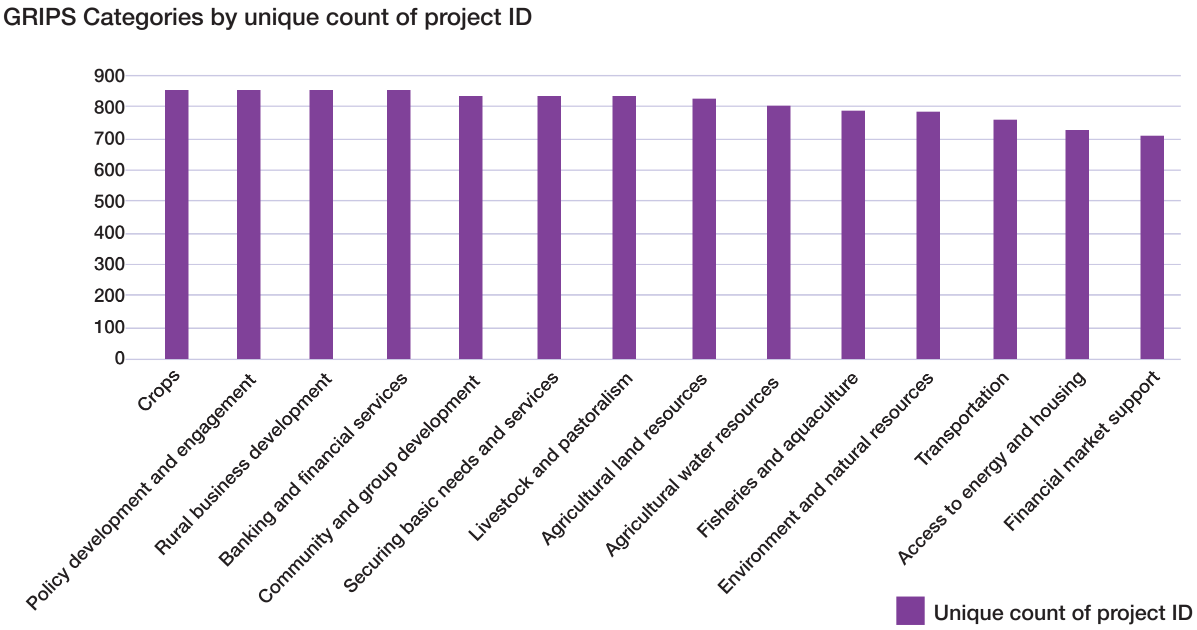


Figure 2 GRIPS Categories by unique project count (n=866 projects).

As mentioned above, the model also detects other project features, such as the different actors targeted by the interventions. Figure 3 shows the overall breakdown of target groups. Farmers, community, and small-scale farmers were the most prevalent groups found.

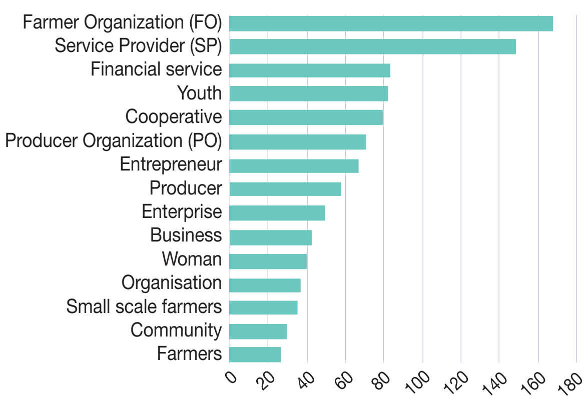


Figure 3 Target groups identified in IFAD projects, by frequency count (n=866 projects)

While the above figures show general results from the project database, the Intervention Dashboard's functionalities allow for the exploration of different areas of interest. It is this searchability that provides focused and granular insights into project activities. Three examples of applications and analysis conducted using the tool are presented below.

CASE STUDY 1. *Value Chain*

Value chain interventions are characterised by initiatives that unite enterprises and agents to bring a product from production to the final consumer. The term "value chain" was searched within the labelled interventions and the dashboard identified 265 projects which contained value chain interventions.

In these results, the GRIPS category mostly associated to the interventions is rural business development, which accounts for 44% of the identified interventions (figure 4). At the sub-category level, which is presented as a heat map that illustrates the frequency of the sub-categories over time, market linkages are the main focus of the projects that contain value chain interventions, as shown in figure 5. It shows the importance of initiatives that strengthen the links between supply chain agents and trading partners. Another relevant subcategory is business development services, which aims to aid enterprises and small business management by offering planning, administrative and financial services or assisting in the development of business plans. The time series analysis presented in the heatmap shows that the focus on value chain as a concept gained popularity in the late 2000s. Figure 6 shows that mainstreaming themes "gender" and "nutrition-sensitive" were the most frequently addressed by these projects.

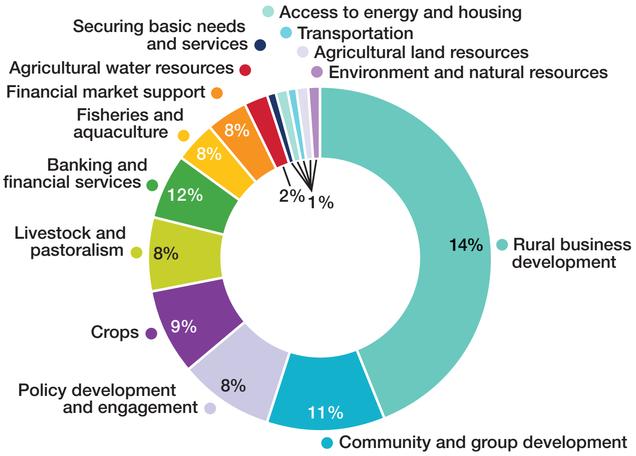


Figure 4 GRIPS categories identified in value chain interventions, by unique project count.

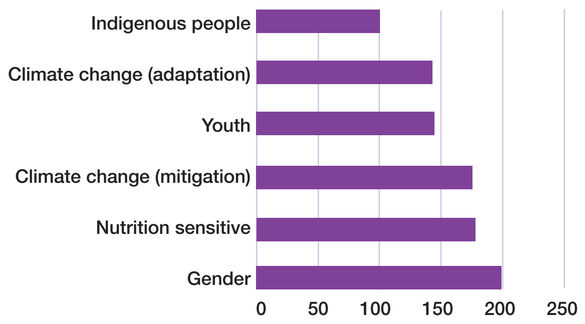


Figure 6 Presence of mainstreaming themes in value chain projects, by unique project count.

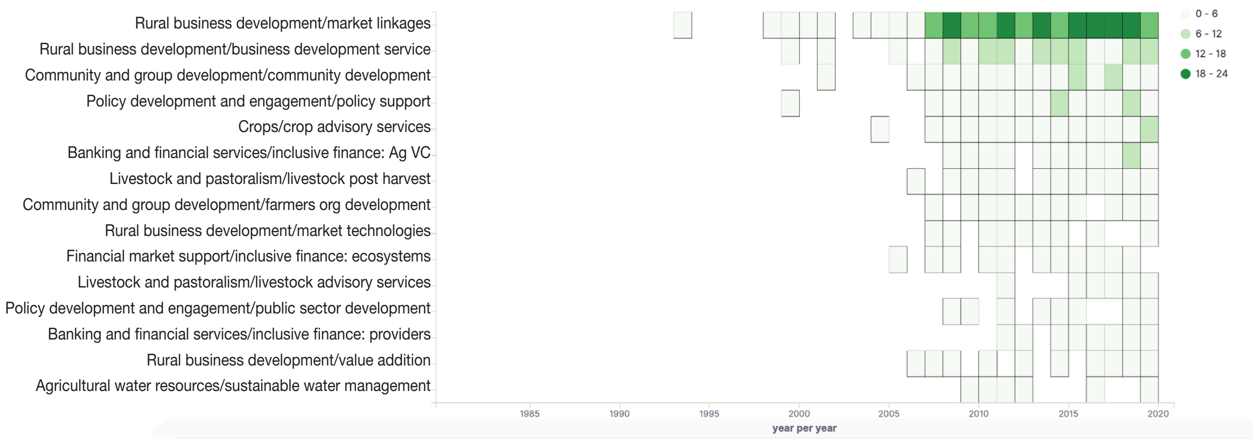


Figure 5 Heatmap of GRIPS subcategories for value chain interventions, by unique project count.

With regards to targeting, the results show a prevalence of value chain interventions aimed at farmers and business enterprises of various shapes and sizes, from small scale farmers to entrepreneurs, cooperatives and farmers' organisations (figure 7). Figure 8 shows that the majority of interventions focused on plant products, with a prevalence of cash crops, vegetables and fruits, whereas animal products detected include poultry, milk and fish. Taking the milk value chain as an example, the model identifies 49 projects mostly located in Rwanda, Bosnia-Herzegovina and Brazil (figure 9), which target primarily farmers and cooperatives, in comparison to the average value chain intervention (figure 10).

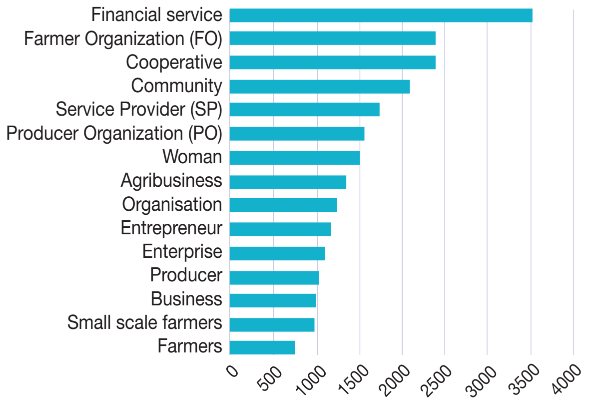


Figure 7 Target groups identified in value chain interventions, by unique project count.

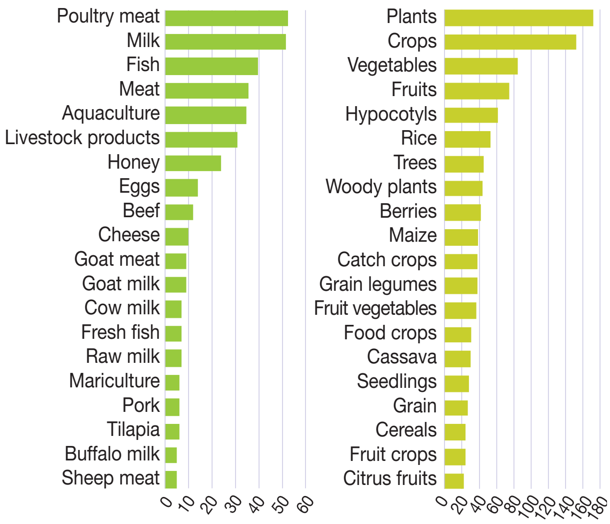


Figure 8 Plant and animal products identified in value chain interventions, by unique project count.

In terms of project effectiveness, an important question is whether the interventions proposed at project design were actually delivered. The dashboard enables cross checking if the interventions found in the PDRs were also present in posterior documents such as the midterm review reports (MTRs) and PCRs for projects where those documents are available. The data shows that, among the value chain projects, 53% had PCR and MTR available. Among those, 45% of interventions detected were present from design to completion.

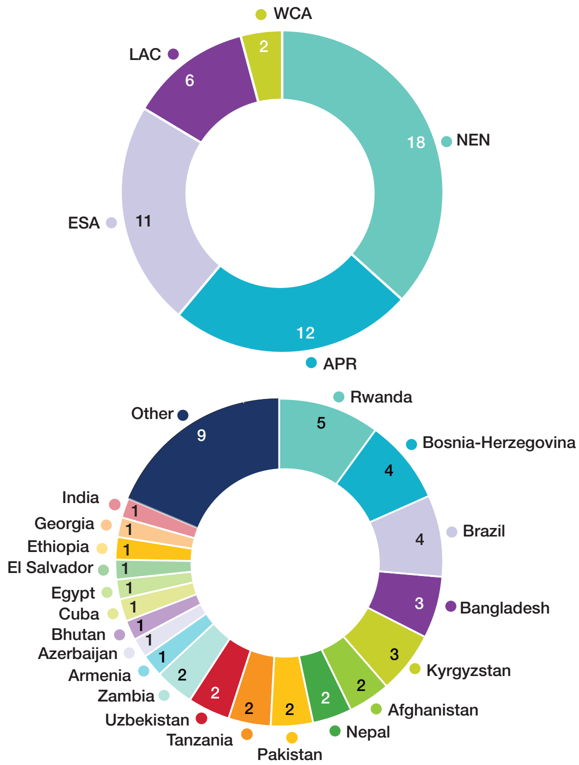


Figure 9 IFAD regions and countries identified in milk value chain, by unique project count.

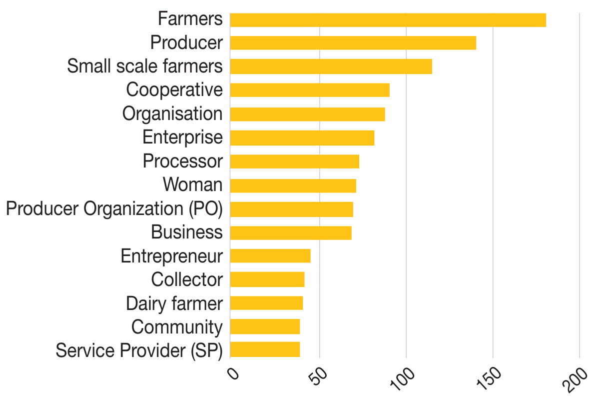


Figure 10 Target groups identified in milk value chain interventions, by intervention count.

CASE STUDY 2. *Rural finance*

Rural finance interventions are present in 279 projects, mostly distributed between APR, ESA and NEN regions (figure 11). Differently from value chain interventions, figure 12 shows that rural finance is not associated with well-defined plant or animal products. For instance, they are related to a smaller range of animal products, dominated by milk and fish. As expected, these interventions primarily target financial institutions at the community and farmer levels (figure 13). According to the subcategory heatmap in figure 14, rural finance interventions have been present in the IFAD portfolio since the 1990s and have been implemented in conjunction with infrastructure and road transport mobility financial support.

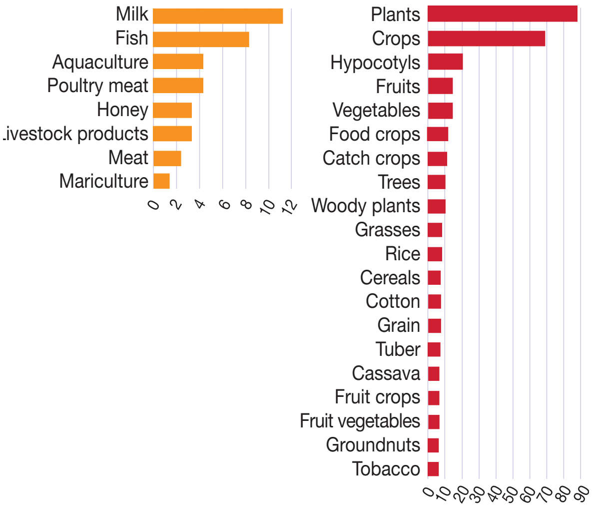


Figure 12 Plant and animal products identified in rural finance interventions, by unique project count.

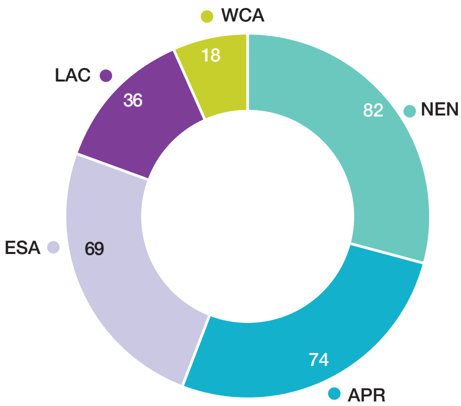


Figure 11 Regions identified in rural finance interventions, by unique project count.

Figure 13 Groups targeted in rural finance interventions, by intervention count.

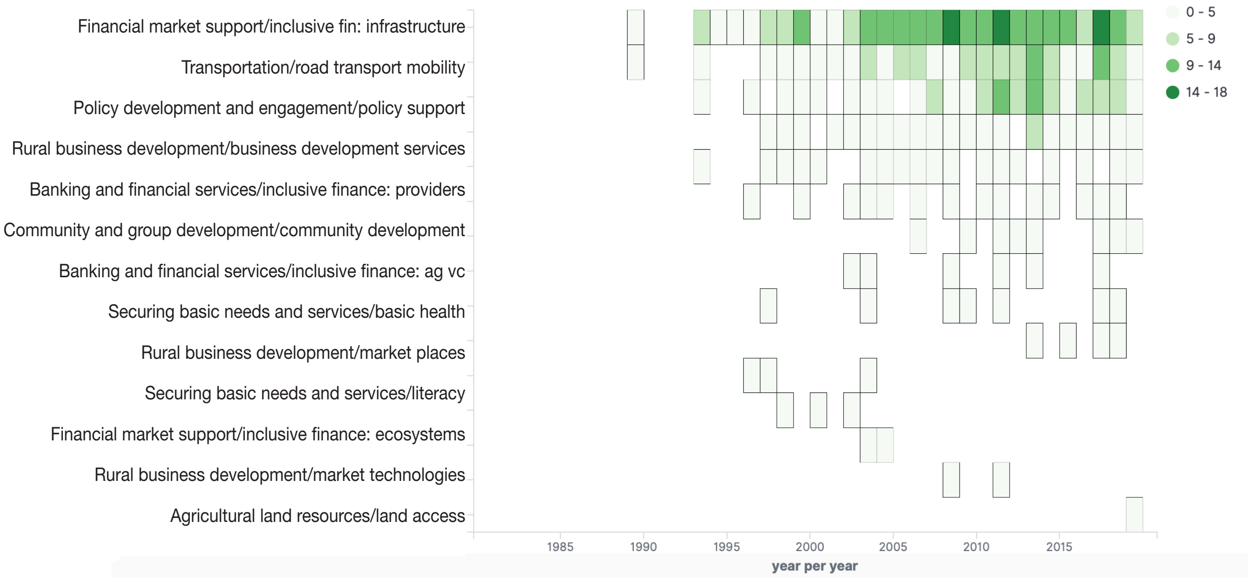


Figure 14 Heatmap for subcategories identified in rural finance interventions, by unique project count.

Figure 15 shows the most prominent expected outcome related to these interventions: the first is livelihood (32%) – which represents income and economic mobility-related impacts – followed by production (20%) and gender empowerment (12%). In figure 16, the major topics detected by the model show that projects involving rural finance are related to policy development around household economic mobility (14%), as well as innovation adoption (12%) and research and development investments (10%).

Lastly, 54.2% of all rural finance interventions found in projects with PDR and PCR in the sample were present from design to completion, suggesting the majority of planned interventions were executed.

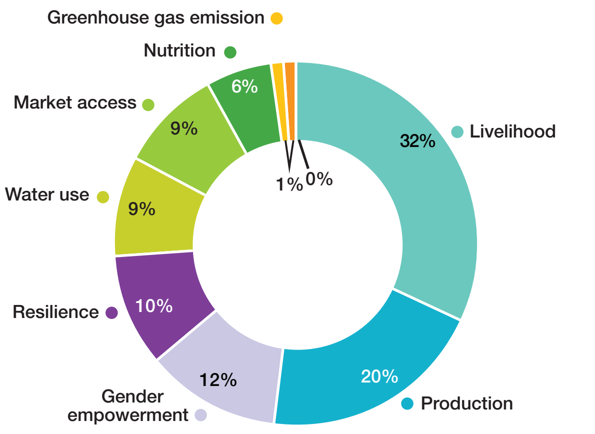


Figure 15 Expected outcomes identified in rural finance interventions, by unique project count.

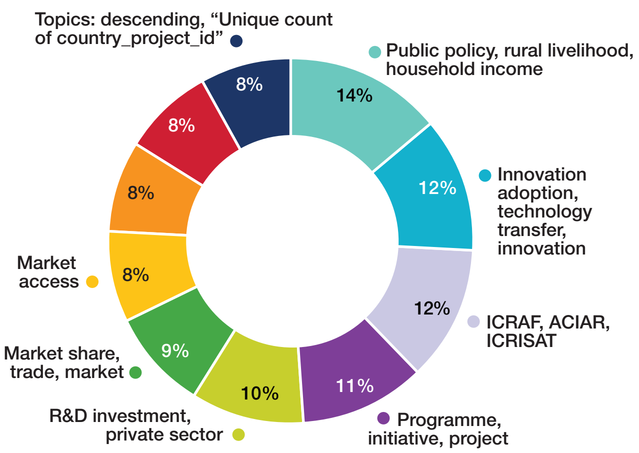


Figure 16 Topics identified in rural finance interventions, by unique project count.

CASE STUDY 3. *Regional Report: Latin American and Caribbean (LAC)*

The dashboard can also provide users with regional overviews. In this case, the data was filtered to look for all projects in LAC. From 144 projects found, the majority have been implemented in Brazil, Bolivia, Honduras, Ecuador and Mexico, as seen in figure 17. The five most frequent interventions detected by project count were financial services, cultivation, animal husbandry, harvesting and institutional strengthening. Contrasting from the theme-specific examples, figure 18 shows that projects in LAC target more organisations, business and communities-level groups than farmers directly.

Figure 17 Countries identified for projects in LAC region, by unique project count.

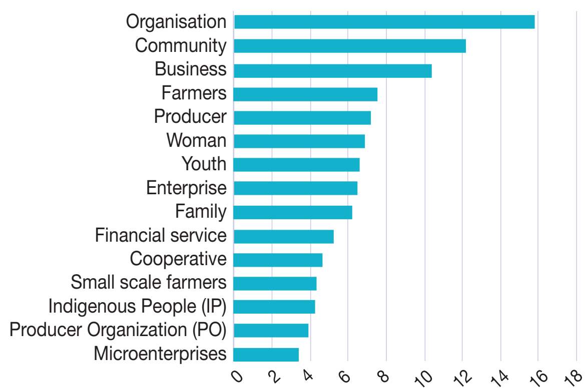


Figure 18 Target groups identified for projects in LAC region, by intervention count.

Regarding the GRIPS classifications, figure 19 shows that LAC projects contain interventions mostly associated with the categories: policy development and engagement, crops and securing basic needs. From those, the related subcategories comprise policy support, crop advisory services, and basic health, respectively. The prevalent plant products are maize and wheat, the two most commonly cultivated grains in the region. The most frequent animal products found are milk and beef, also key productive sectors, as shown in figure 20.

Lastly, the most frequent outcomes expected from project interventions in the region are related to production, livelihood, water use and gender empowerment (figure 21).

Figure 19 Heatmap for GRIPS subcategories identified for projects in LAC region, by unique project count.

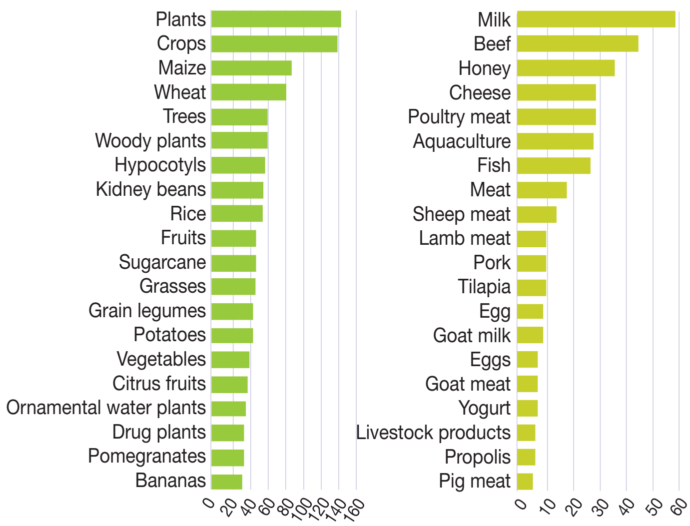


Figure 20 Plant and animal products identified for projects in LAC region, by unique project count.

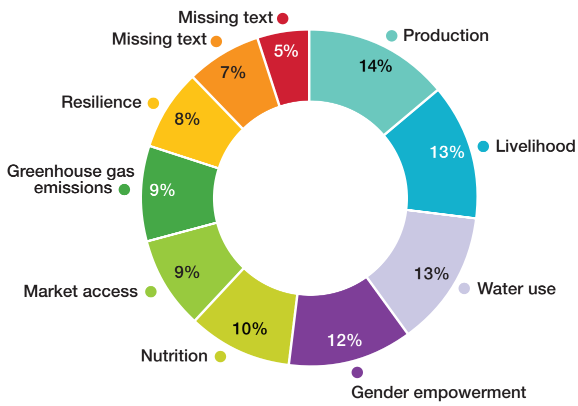


Figure 21 Expected outcomes identified for projects in LAC region, by unique project count.

Feedback from dashboard users

Product validation sessions were conducted with key IFAD staff in order to present the dashboard and elicit feedback from the user end. Users also had access to a detailed manual and an introductory video. The overall response to the tool was very positive, with users recognizing the utility of the dashboard for their own data needs.

The usefulness of granular results was highlighted. One reviewer recalled having been asked many times, prior to using the dashboard, if they "could determine which projects focused on a specific crop and it was impossible to find a reliable answer. This is one of the strong points and added-value of this system!" The possibility of carrying out thematic searches to produce case studies was also noted.

The feedback made it clear that more training is needed, so that users both understand how the data is analysed and are able to make the best use of the tool. Integration with ORMS and other systems were suggested as possible next steps so that users could link directly to the documentation analysed.

The choice of project documentation from which the bulk of the analysis was done was also considered. In phase 1, focus was placed on PDRs, as these were standardised documents written by IFAD staff. However, users felt that a key piece of information missing from the dashboard was whether the identified interventions had indeed been executed or were just mentioned at project design. In response to this, a new pilot feature was added which looks for the interventions detected in a PDR and PCR/MTR (where available) of that same project. This is an initial approximation of the extent to which an intervention was present from design to completion.

Moving forward, the model and underlying algorithms would need to be improved in the following way, as also indicated by the users:

1) To date, the model only processes one document at a time per project. Therefore, algorithmic modifications are required to allow data processing and aggregation of multiple sets of documents per project - PDR, MTR, project supervision reports (PSR), and PCR, as well as integration of cost-data for triangulation purposes.
2) The model only considers PDRs at the moment. Additional programming is needed to allow triangulation between PDRs (which state foreseen interventions by the project) and PCRs (which contains implemented interventions) to really gauge whether an intervention was implemented on the ground, and to take into account the reality of adaptive implementation. One avenue of future model improvements would be to allow the model to distinguish prospective language (can, will and would, should and shall) which reflect a future intention, to language related to realized objectives (did, done, benefitted from) and let the classification algorithms be driven by "realized language" as opposed to prospective statements.
3) Cost data will need to be integrated in the subsequent Phase of this project.

Lessons Learned Application: identifying topics from lessons learned at project completion

This workstream leveraged PCRs gathered in phase 1 to support the project cycle by harnessing learning from the documentation through the development of a taxonomy of lessons learned. Classifying and grouping the lessons learned from PCRs according to a set of characteristics can facilitate knowledge management and referencing for future IFAD projects. For such cases, topic modelling techniques are useful to identify relevant themes from text samples and create those groups. The chosen approach used a combination of unsupervised and semi-supervised machine learning algorithms.

Text from the Lessons Learned section of 469 PCRs were extracted, with sentiment analysis and topic modelling approaches applied for the analysis. The final output is a user-friendly application built in the "R" programming language, the Lessons Learned Web App, which enables searching and analysing through the identified topics.

CASE STUDY 4. *Using the web app: learning from stages in the project cycle*

The starting point in this example are three topics, each related to a stage of the project cycle: Design (Topic 1), Delivery (Topic 2), Outcome (Topic 3); while Topic 4 groups topics which did not correspond to the other three. The Lessons Learned web app allows the visualization of the most common terms corresponding to each topic, which give an indication of what the lessons learned from those themes are about. Specifically, figure 22 shows that lessons learned related to project design are often about project development, support and management, whereas outcomes-related lessons focus more on how the intervention may have affected communities, partners or farmers. Analysis of the topics against other variables is also possible like, for instance, the correlation between topics and project regions. Figure 23 shows the probability of topic occurrence according to the project region. It appears that much of the learning from LAC and WCA projects is related to project design, whereas lessons learned from APR, ESA and NEN are distributed among delivery and outcomes.

From the overall results, it is also possible to unpack specific terms. As an example, figure 24 shows the presence of the word "participatory" in the three topics. Unsurprisingly, this term is most frequent in topic 2, which reflects IFAD's approach to project implementation. A final resource is the possibility to read an extract of the learning where the selected word appears, classified by project ID and topic cluster. As such, the app enables the users to filter through terms and extract the specific passages where the topic of interest is present.

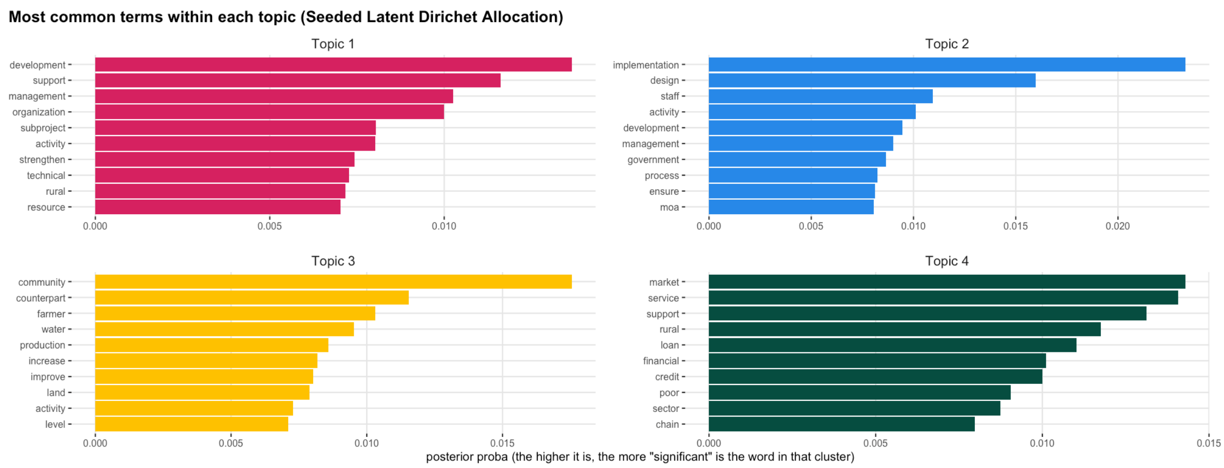


Figure 22 Most common terms within each topic for Topic 1 = Design; Topic 2 = Delivery, Topic 3 = Outcome.

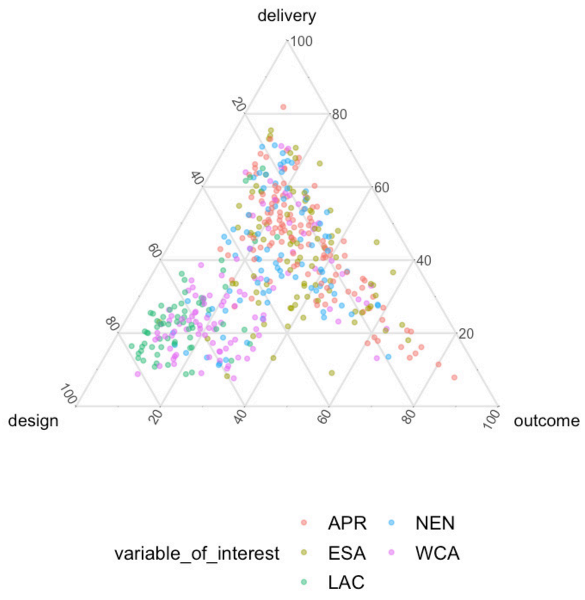


Figure 23 Correlation of topics with project regions.

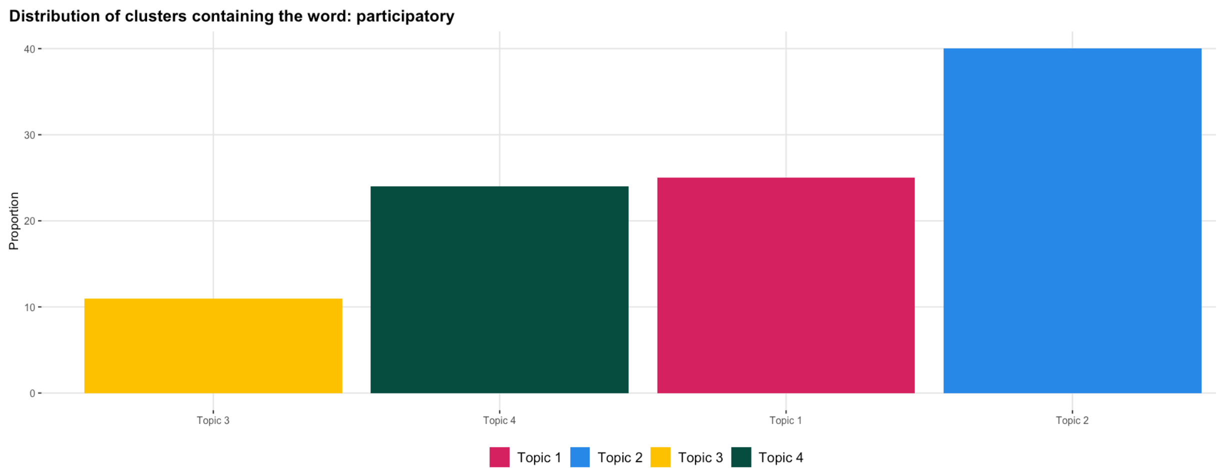


Figure 24 Distribution of clusters containing the word "participatory" for Topic 1 = Design; Topic 2 = Delivery, Topic 3 = Outcome.

	ID	posterior	cluster	extract
57	1100001061	0.286301369863014	Topic 1	was exactly the opposite of what the lesson indicatedseveral drawbacks follow from this firstly the participatory approach which was abandoned for the reasons presented previously became all the more d
58	1100001061	0.286301369863014	Topic 1	ial details of implementation and made standardized project evaluation impossiblesecondly while the participatory development approach was soon abandoned in the sgrdp it remains unclear whether it was
59	1100001061	0.286301369863014	Topic 1	mselves for concrete results such experience may lead us to seriously question the relevance of the participatory approach in certain contexts indeed in very poor communities a consensus on very basic
60	1100001061	0.286301369863014	Topic 1	rities may emerge only at higher levels of social developmenthowever the apparent redundancy of the participatory approach may in fact be a symptom of the fact that there was not enough participatory a
61	1100001061	0.286301369863014	Topic 1	ndancy of the participatory approach may in fact be a symptom of the fact that there was not enough participatory approach indeed the experience of the sgrdp shows thatrepublic of yemen � southern gove

Figure 25 Extract of the documents where the selected word appears.

Trends Analysis: uncovering historical IFAD investments in strategic themes

In phase 1, a descriptive text mining technique was applied to project documents with the objective of identifying how projects reported against two aspects: (i) IFAD's mainstreaming themes; and (ii) the Sustainable Development Goals. The key element of this model was the development of custom taxonomies, against which project documents can be mapped. In the second phase the team validated the text mining model and expanded the analysis to SDG targets, especially relating to SDGs 1 and 2. We also employed the same approach to explore frameworks for food systems and ICT4D, consolidating the applicability of machine learning to understand historical trends about IFAD's activities regarding strategic themes.

In total, 1,769 documents (PDRs, MTRs and PCRs) for 849 projects with approval dates ranging from 1981 to 2019 were analysed through the development of multi-lingual algorithms. Table 2 reports statistics for the number of projects by sector. The majority are classified as rural development (35%) and agricultural development (29%).

Table 2 Number of projects per IFAD sector

Sector	Number	Percentage
Rural Development	296	34.86
Agricultural Development	250	29.45
Credit and Financial Services	127	14.96
Research/Extension/Training	46	5.42
Irrigation	43	5.06
Marketing/Storage/Processing	36	4.24
Livestock	30	3.53
Fisheries	16	1.88
Other	5	0.59
Total	**849**	**100.00**

The text was extracted from the reports to create a merged corpus containing more than 138,000 pages of text and a "term document matrix", which identified more than 170,000 unique terms for all languages, distributed across the documents.

Uncovering IFAD investments against the Sustainable Development Goals

Expanding on the analysis of SDG coverage in project documents performed in phase 1, this iteration considered the 169 specific targets devised for achieving each SDG[6]. The advantage of this approach is the possibility to aggregate findings according to the hierarchical structure of the SDGs. Following the technique already established in phase 1, taxonomies were created for each target through a word expansion model. Moreover, a more refined text extraction procedure was introduced which focused on parts of the documents that contained a reference to the term "goal" and similar words identified through automated word expansion (e.g. aim, objective, strategy, etc.).

In order to understand the most relevant terms within the IFAD documentation, a measure of word significance for the text extracted was applied. Analysis of this taxonomy against the terms associated to the SDG targets is then able to measure the presence of SDG-related content in the reports.

Figure 26 shows the overall distribution of the SDGs across IFAD documents. While this shows a heterogeneous distribution, higher prevalence has been found for IFAD's key SDGs: SDG 1 – No Poverty and SDG 2 – Zero hunger. Three other SDGs show strong presence: SDG 8 – Decent work and economic growth, SDG 12 – Responsible consumption and production, and SDG 17 – Partnerships.

6 https://www.un.org/sustainabledevelopment/

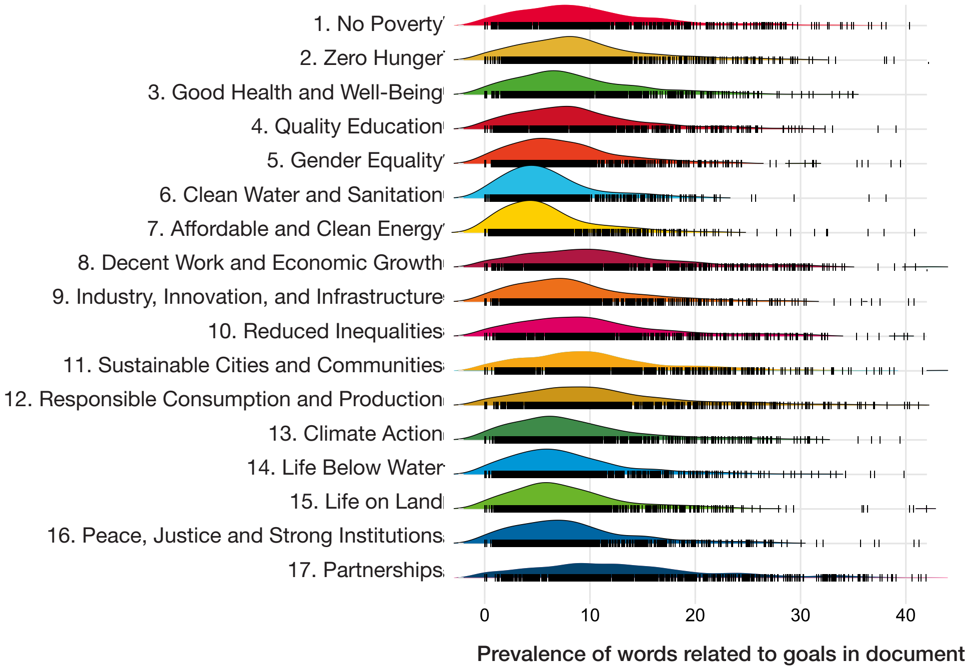


Figure 26 Distribution of SDGs in IFAD documents (n=1769).

A degree of heterogeneity is also observed comparing the average prevalence of SDGs at the design stage of projects (as present in PDRs) and at the completion stage (as present in PCRs). With regard to SDG 17 – Partnerships, PCRs contained more tt associated to this goal than their respective PDRs, indicating that projects increase their focus on partnership building over the project cycle. Likewise, gender-related content (SDG 5) is more prevalent in PCRs. Conversely, content associated to SDG 10 (Reduced Inequalities) was more prevalent at PDR stage.

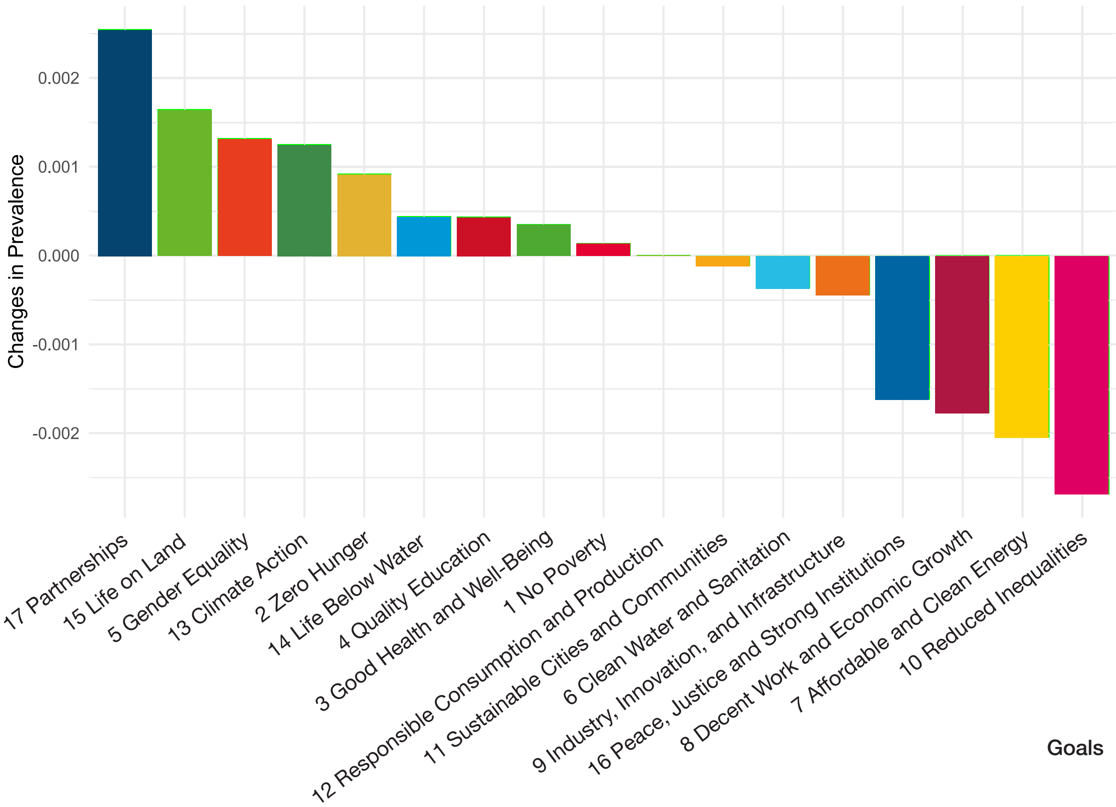


Figure 27 Change in prevalence of Goals from PDRs to PCRs (n=469).

To assess trends over time, the themes were analysed against the 11 replenishment periods. An increase in prevalence was observed for all SDGs. Figures 28 and 29 present the results for SDGs 1 and 2. The colours represent quartiles, and it is possible to see a consistent increase in the median over the course of replenishments.

Still focusing on SDGs 1 and 2, the main ones related to IFAD's strategic objectives, figure 30 presents the distribution of specific targets in the documentation. Regarding SDGs 1, it shows a higher correlation for target 1.4. For SDG 2, targets 2.3, 2.4 and 2.5 were the most prevalent (see here for a list of SDGs and their related targets: https://sdgs.un.org/goals).

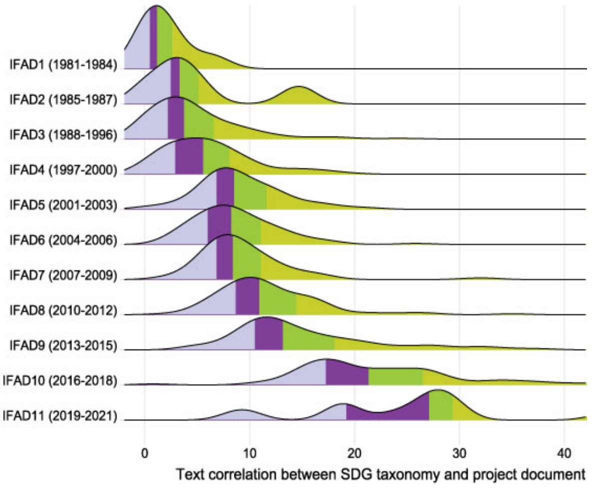


Figure 28 Distribution of Goal 1 by IFAD replenishments.

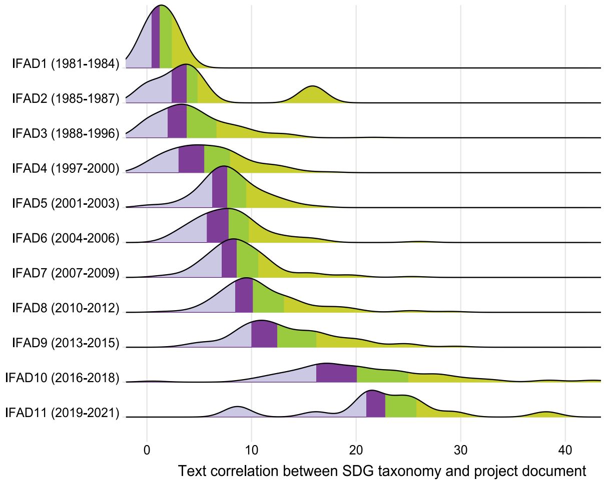


Figure 29 Distribution of Goal 2 by IFAD replenishments.

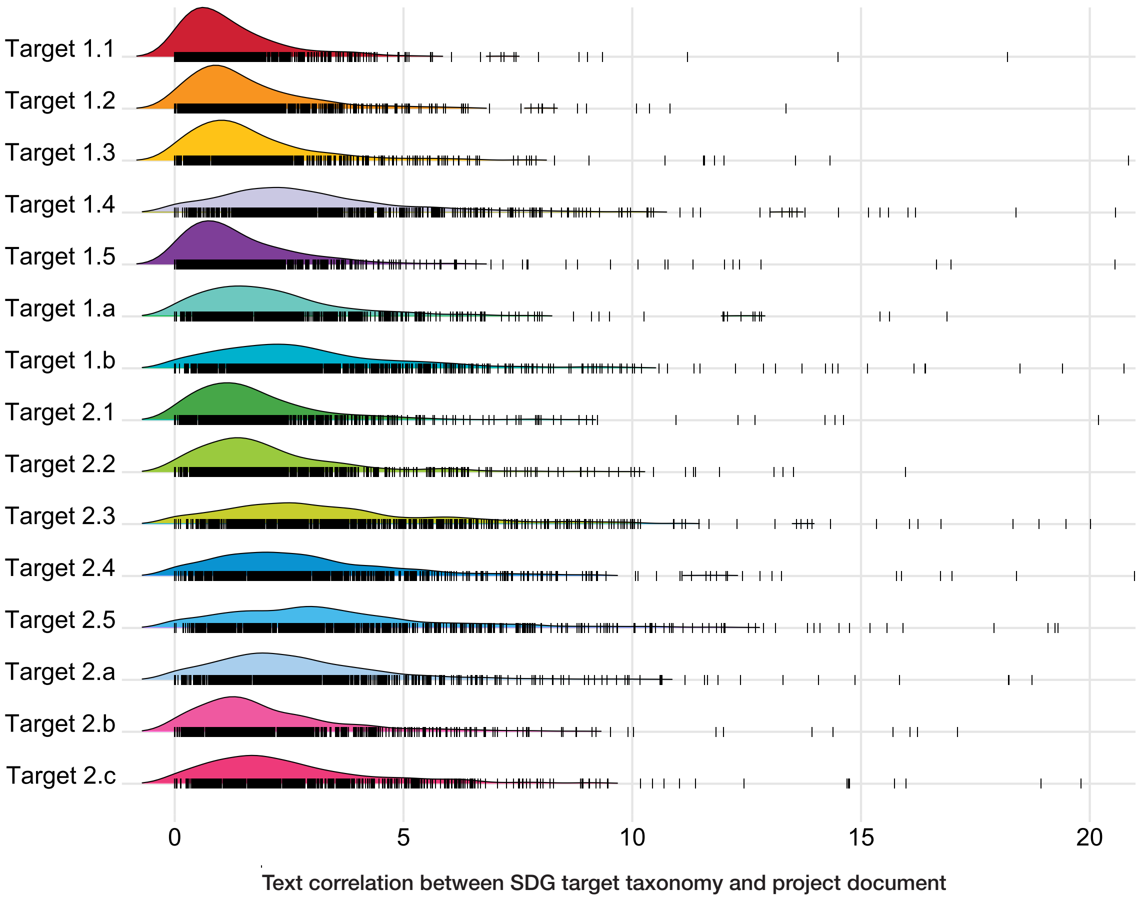


Figure 30 Distribution of targets for SDGs 1 and 2 in project documentation.

Figure 31 is a heat map of the distribution over the replenishment periods, which supports the overall analysis that reporting against the SDGs has increased over time. The darker shades in the map show a greater prevalence of specific targets in the latter replenishment periods. The targets presenting the most significant increased prevalence over time are the same as those with the largest correlations, namely SDG targets 1.4, 2.3, 2.4 and 2.5.

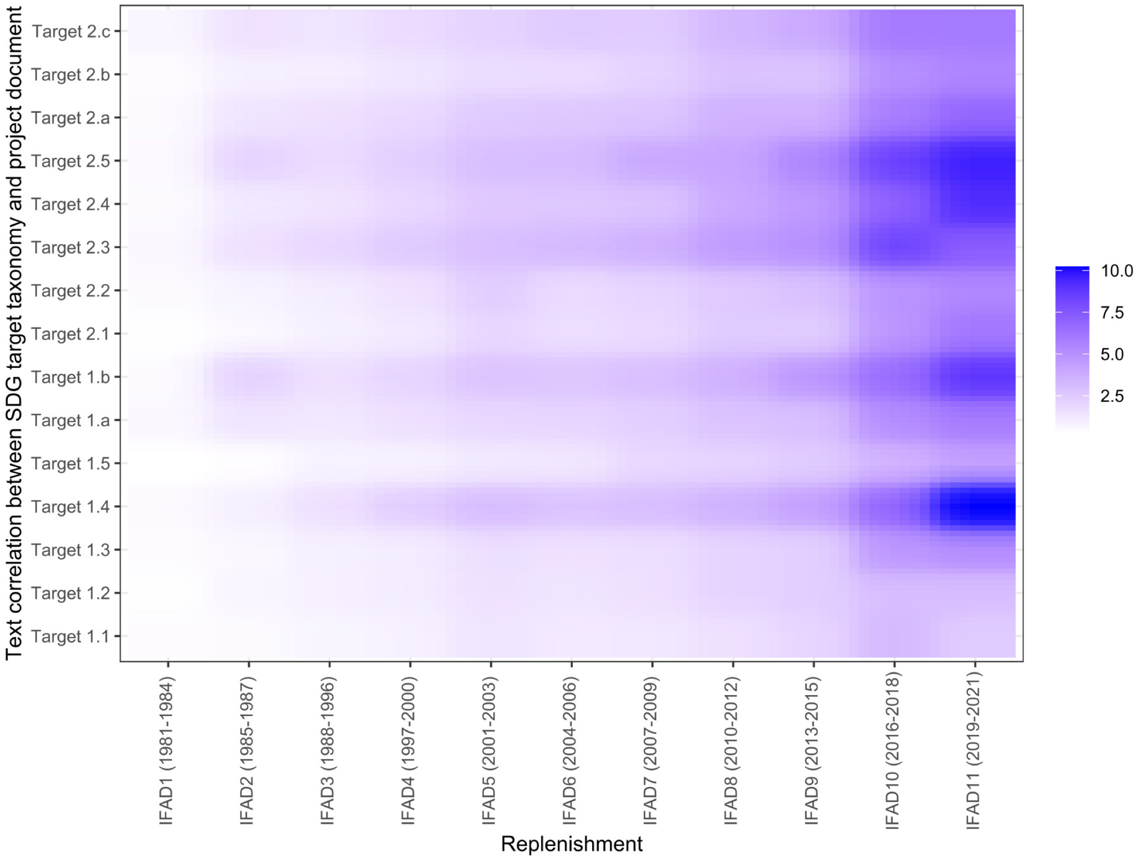


Figure 31 Heatmap of the distribution of SDG targets for Goals 1 and 2, by replenishment period.

Understanding IFAD investments in food systems

A background paper was submitted to the IFAD Rural Development Report (RDR) 2021 on Food System Transformations, which aimed to provide evidence on the dynamics between food system dimensions within IFAD-funded projects. By employing the aforementioned text mining approach to project documentation, complemented by other machine learning techniques, it uncovered relevant trends in IFAD's investment portfolio related to food system types, components and outcomes. Questions about the presence of food system dimensions in project documentation and the interaction between these were explored through supervised text mining, network analysis and a LASSO regression analysis. The study considered 1,679 project documents collected from 849 projects in IFAD's investment portfolio since 1981 (PCRs, MTRs and PCRs).

The starting point was the development of a taxonomy based on IFAD's analytical framework for the analysis of food systems transformations and the literature used to establish it.[7] Figure 32 illustrates the framework, with the three dimensions and their analytical sub-categories, from where the taxonomy was developed. Based on definitions of these analytical categories available in the literature, key words for each concept were identified and manually extracted in order to establish a terminology from which to map the data against.

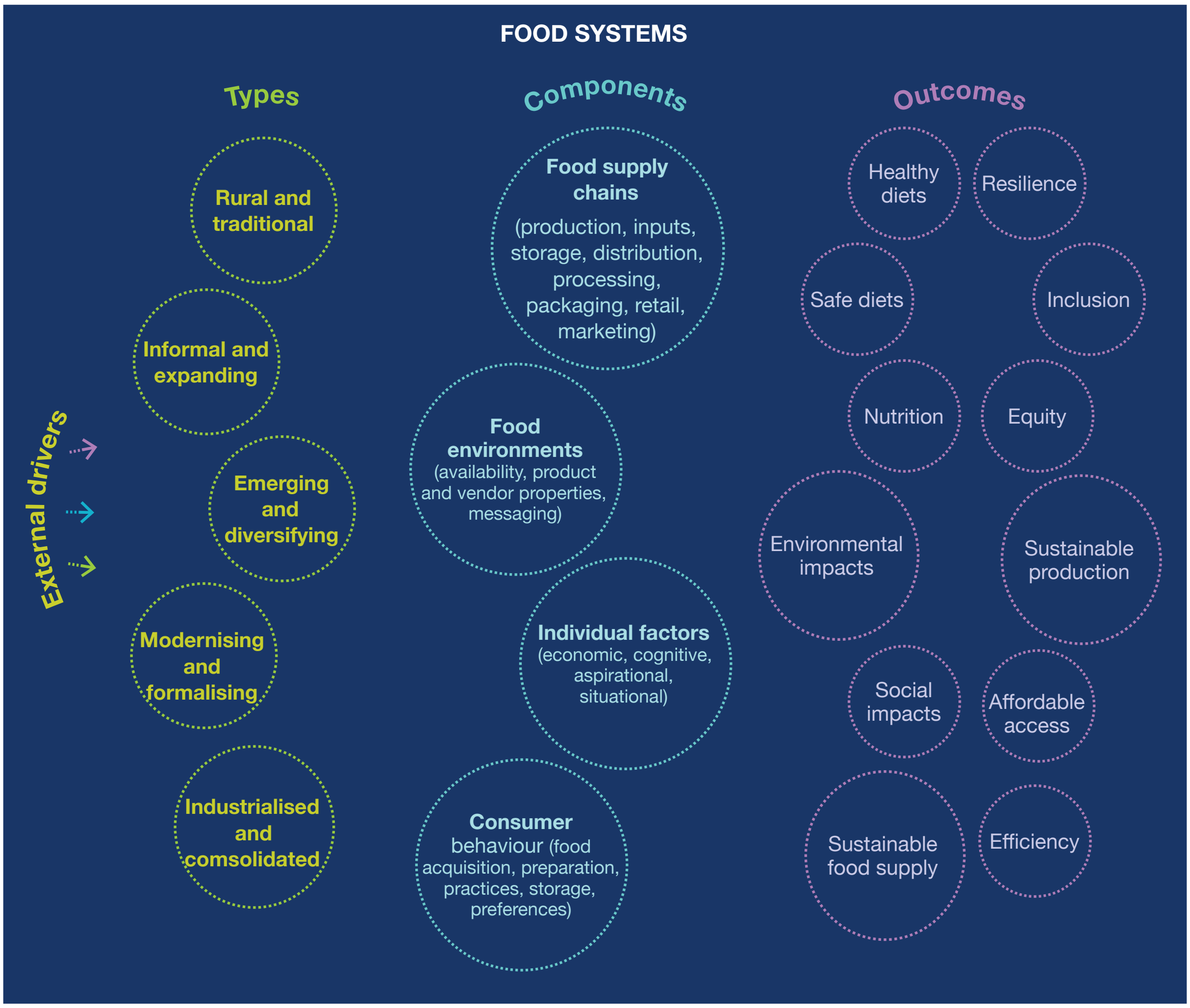


Figure 32 Food systems framework for supervised text mining of IFAD project documentation.

7 GLOPAN (2016) Food systems and diets: Facing the challenges of the 21st century; IFAD (Forthcoming). IFAD Rural Development Report 2021 on Food System Transformations; HLPE (2017) Nutrition and food systems. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security.

Overall, descriptive text mining uncovered an upward trend in reporting against food system dimensions over the time interval considered in the analysis. The first explorative exercise was to identify the share of words associated to food system archetypes, components and outcomes in each document. Figure 33 shows the overall distributions of the food systems archetypes. It is possible to identify a prevalence of both "rural and traditional" and "modernising and formalising" systems, as these were the types with the widest distribution of percentages across the portfolio. The other three, "informal and expanding", "industrialised and consolidated", and "emerging and diversifying" are still well spread, but their share of words is less distributed. These results are not surprising given IFAD's focus not only on small-scale rural producers, who may still carry out traditional farming practices, but also on improving their livelihoods through increased productivity and market access.

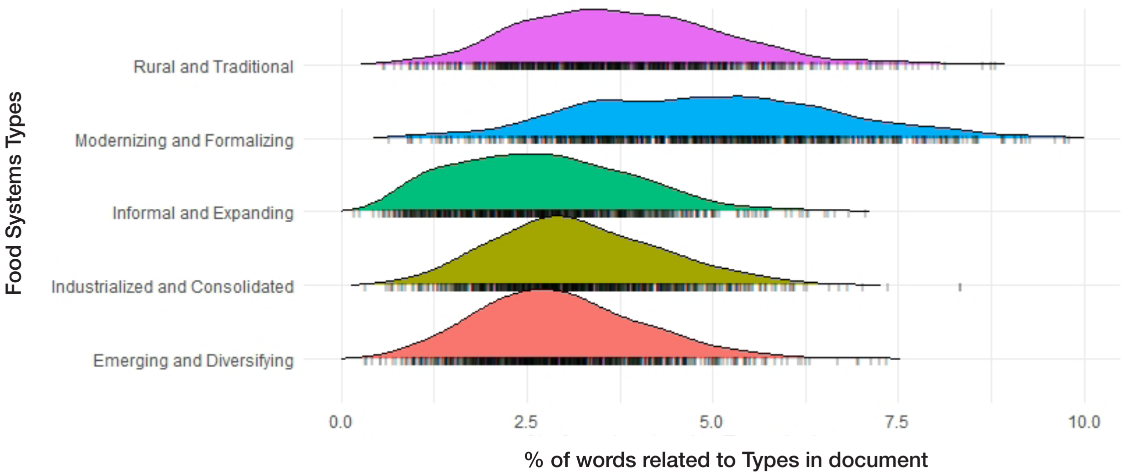


Figure 33 Distribution of food systems archetypes in project documentation. Share of words associated to food system archetypes present in project documentation. Key terms collected manually and expanded by word2vec (1769 documents for 849 projects analysed).

As discussed in IFAD's RDR analytical framework, *"several adjustments in food systems take place simultaneously with heterogeneous rhythm and speed and various food systems might therefore co-exist."* (RDR, 2020:2). This aspect is reflected in the project documentation, which showed a high level of co-occurrence between the different food system archetypes, as illustrated by figure 34. For instance, projects that contain words related to "rural and traditional" food systems are also highly likely to contain words related to "industrialised and consolidated" food systems. These positive correlations between food system types denotes that distinct types are complementary to each other, which can be interpreted as an indication that multiple food systems are recognised in IFAD interventions. Further analysis of how these evolve over the project cycle will be shown below.

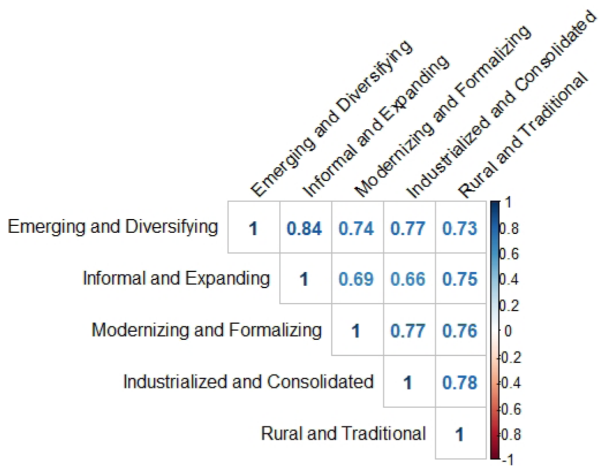


Figure 34 Correlation between food systems archetypes.

Figure 35 presents the distributions of the food systems components. As expected, the category "food supply chains" shows the largest distribution of percentages across the portfolio, followed by terminology related to individual factors – such as income, purchasing power, food affordability and product properties. This overall prevalence can be explained by the close association of food supply chain's categories to IFAD's actual project components and interventions, as projects focus their activities in agricultural production, storage, distribution and value addition for marketing purposes.

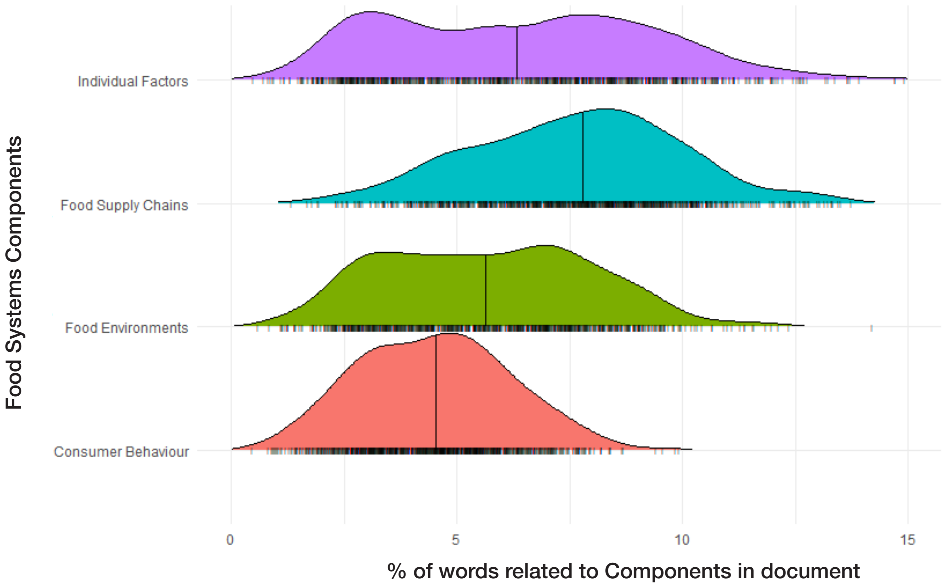


Figure 35 Distribution of food systems components in project documentation. Share of words associated to food system components present in project documentation. Key terms collected manually and expanded by word2vec (1769 documents for 849 projects analysed).

Again, the interactions and synergies of food system elements are visible through the correlations between component categories presented in figure 36, which show, for instance, that projects with a focus on food supply chains are also likely to cover aspects of food environments such as diversity and quality of food supply, prices, and characteristics of market opportunities.

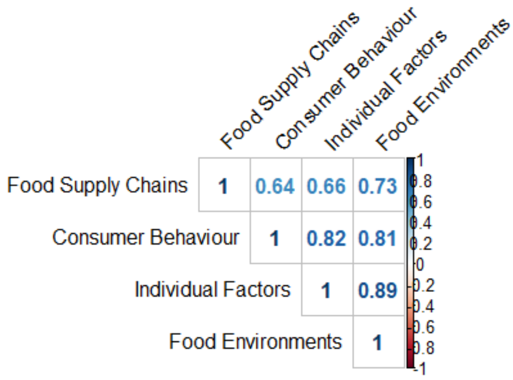


Figure 36 Correlation between food s ystems components.

The presence of food system dimensions in IFAD project documentation can be further visualised through the network graph in figure 37 which shows the connections between projects and the three layers of food system dimensions. The force-directed algorithm used in constructing the network displays the spatialization of nodes, which maps the proximity and the authority of categories in relation to each other. A modularity algorithm was applied to identify clusters, which are coloured accordingly. The dimensions are represented by the colours of the labels.

The resulting network indicates that the component "food supply chains" is the most prevalent category across the documentation. The modularity analysis also identifies a few clusters: the largest one is composed of "food supply chains" with all food system types and the outcomes "sustainable food supply", "sustainable production", "healthy diets" and "nutrition". The Lasso analysis corroborates these relationships, as will be shown next.

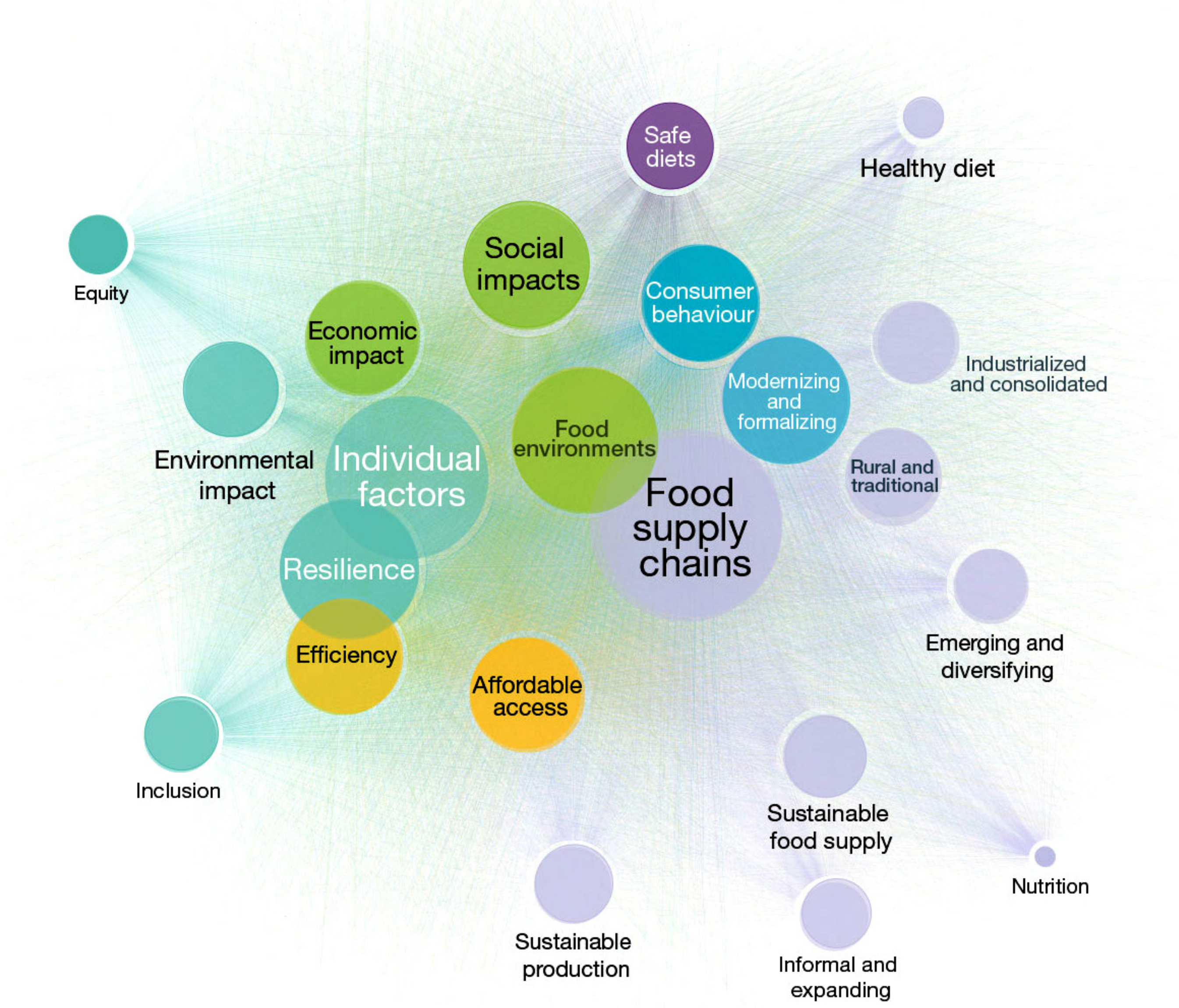


Figure 37 Network of Food Systems dimensions detected in IFAD documentation. Parameters: force-directed graph, with node size partitioned as Weighed in-Degree, coloured by modularity class. Labels coloured according to dimensions. 2324 nodes (project documents + categories), 45149 edges (weighed by share of words).

Figure 38 presents the average changes in prevalence of archetypes between PDRs and PCRs. These changes have been obtained by the difference between the presence of archetypes in a PCR and a PDR. Positive values indicate that there has been an increase in the presence of the archetype in the PCR compared with PDR (across the project duration), while the opposite is true for negative values. It can be noted that the presence of modernizing and formalizing, emerging and diversifying, and industrialized and consolidated increases across the project duration, while the presence of rural and traditional and informal and expanding decreases. The graph is an indication of the expected theory of change of IFAD projects, which are framed and justified within an initial context of traditional or informal systems, but which aim to lift target areas and/or beneficiaries up to a modernizing and formalizing system through increased productivity, enhanced market linkages, and more complex, regulated food supply chains. The higher magnitude in the increase of modernizing and formalizing archetype provides additional support to this interpretation.

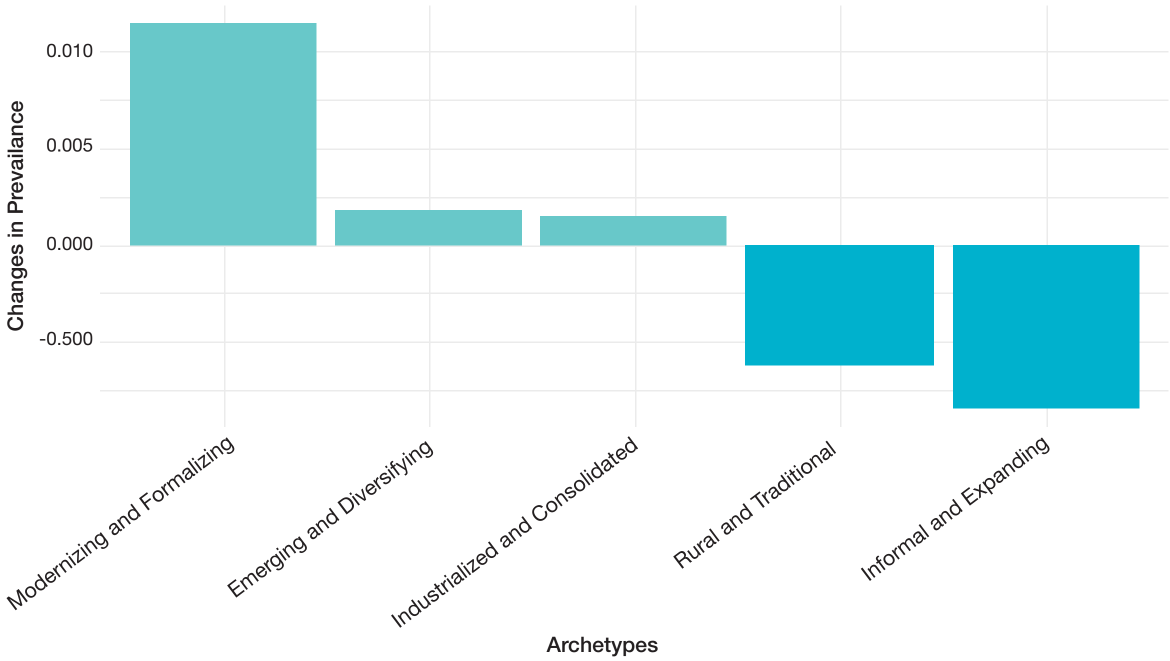


Figure 38 Change in average prevalence of archetypes between PDR and PCR.

Figure 39 shows changes in the average prevalence of components from PDRs to PCRs. As in figure 38, positive values in indicate that there has been an increase in the presence of a component in the PCR, compared with PDR. The increase in individual factors, consumer behaviour, and food environments, in contrast to the decrease in food supply chains could once again reflect the project cycle as represented in the documentation. As PDRs detail the interventions intended for a particular project, and since food supply chain concepts are closely linked to IFAD's scope of interventions, they are naturally more present in this type of report. On the other hand, PCRs reflect on a project's achievements of its development objectives and any implications to a broader social and economic context, thus opening up space for greater dialogue with the other components.

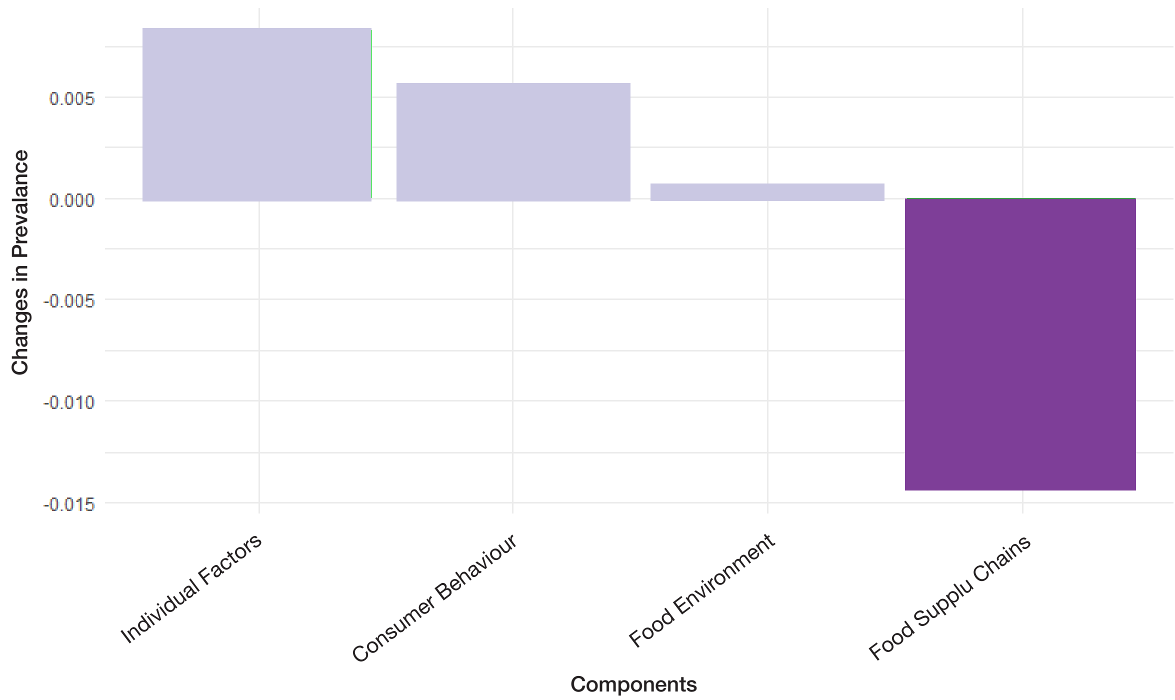


Figure 39 Change in average prevalence of food system components between PDR and PCR.

Regarding livelihood outcomes, LASSO regressions were performed using the share of words identified for the four components (individual factors, consumer behaviour, food environments, and food supply chains) as predictors, and the share of words identified for the 13 outcomes (affordable access, economic impact, efficiency, environmental impact, equity, healthy diet, inclusion, nutrition, resilience, safe diets, social impacts, sustainable food supply, and sustainable production) as the dependent variables in 13 different LASSO regressions specifications. All the models have been processed by cross-validated resampling (10-fold, repeated 5 times) and the final lambda[8] used in the model is selected by the root means square error (RMSE) out of the sample (i.e. the best predictive model out of the sample, in the spirit of ML).

The feature importance of components in predicting each outcome are presented in Table 3 and support the connections highlighted in the network analysis. "Food supply chain" is the most important feature in predicting: nutrition, sustainable food supply, and sustainable production; but not in explaining affordable access, healthy diet and inclusion. "Individual factors" is the most important feature in predicting affordable access, economic impact, environmental impact, equity, inclusion, resilience, and social impacts, but not nutrition, safe diets and sustainable food supply. "Food environment" is the most important feature in predicting efficiency, but does not explain social impacts. Lastly, "consumer behaviour" is the most important feature in predicting healthy diet and safe diets, but not economic impact, efficiency, environmental impact, equity, resilience and sustainable production.

8 Lambda is a specific tuning parameter in ML that is used to assess the performance of statistical algorithms and control their behaviour. The basic intuition is to choose it sensibly in order to get the best model.

Table 3 feature importance for component in predicting outcomes estimated with LASSO models (1769 documents for 849 projects analysed).

	Individual Factors	Consumer Behaviour	Food Environments	Food Supply Chains
Affordable Access	100,00	5,81	46,24	0,00
Economic Impact	100,00	0,00	43,41	43,06
Efficiency	77,97	0,00	100,00	32,80
Environmental Impact	100,00	0,00	73,17	16,88
Equity	100,00	0,00	4,63	21,50
Healthy Diet	20,26	100,00	2,06	0,00
Inclusion	100,00	16,23	67,16	0,00
Nutrition	0,00	76,12	19,66	100,00
Resilience	100,00	0,00	31,14	3,14
Safe Diets	0,00	100,00	37,30	31,27
Social Impacts	0,00	100,00	37,30	31,27
Sustainable Food Supply	0,00	9,14	23,13	100,00
Sustainable Production	34,98	0,00	0,00	100,00

The objective of this study was to identify the dynamics of food systems archetypes, components and outcomes in IFAD project documentation as evidentiary sources of the project cycle. Through the ML techniques, insights were drawn that shed light into IFAD's activities in this subject by providing a historical overview of food system dynamics over 40 years.

First, the positive correlations between food system archetypes reflect the complexity of food systems, where multiple food systems coexist depending on factors such as regional and productive differences. Likewise, components interact with each other, though in IFAD's context, the food supply chain component showed a strong prevalence, as the activities involved in taking food from the farm to the table comprise the bulk of interventions implemented through projects.

Secondly, the changes in average prevalence of food systems archetypes, from rural and informal systems at project design to modernizing and formalizing at project completion, indicate that the impact pathways of IFAD projects aim to improve the livelihoods of rural smallholders through integration into longer supply chains and more complex systems. In addition, the LASSO analysis enabled interpretation of the relationship between components and livelihoods outcomes.

In sum, this study shows the interlinkages between food system dimensions as represented in IFAD project documentation, with many possible pathways between the various elements. While IFAD activity is centred on the food supply chain component, the significance of individual factors for various outcomes highlights the need to strategically combine food supply interventions with consumption-driven initiatives.

Presence of ICT4D in projects

This part of the project explored the extent of ICT4D terminology present in IFAD documentation. The starting points were the terms defined in the IFAD's ICT4D glossary, as well as a draft proposal for ICT4D solutions proposed by APR. Note that this is work in progress and we can update it once we have the final taxonomy. Figure 40 shows a word cloud that contains the frequency of stemmed words in the applied taxonomy.

This taxonomy was then mapped against the project documentation previously described. Figure 41 shows the distribution and the average presence of ICT4D in IFAD documents, by replenishment period. Overall a wide increase can be observed from IFAD1 (4.1% in average) to IFAD11 (9.9%), with a maximum presence (13.6%) in IFAD7.

Figure 40 Word cloud of all (stemmed) words used in the ICT4d taxonomy.

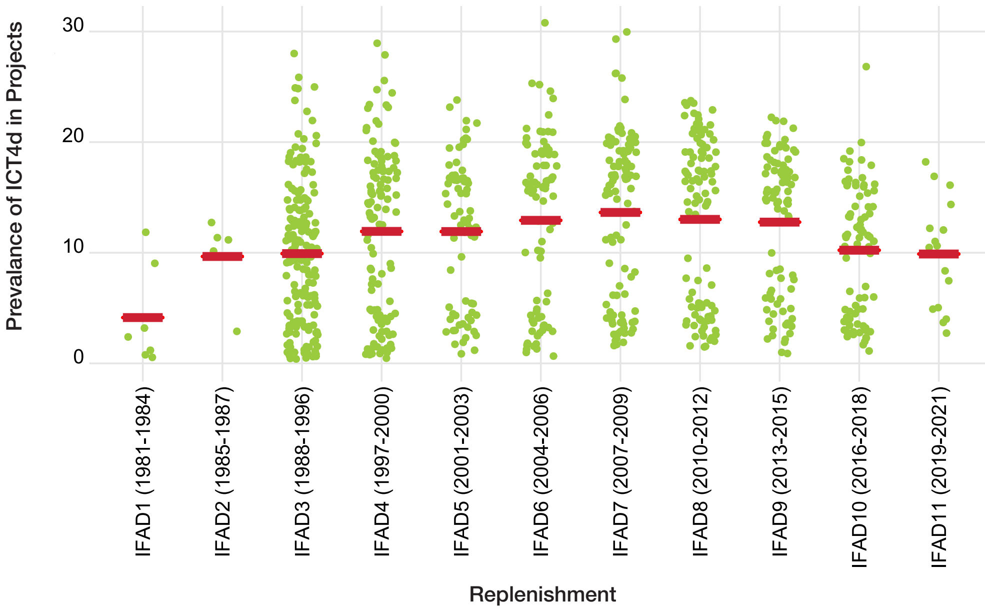


Figure 41 Distribution of ICT4D taxonomy in IFAD project documentation, by replenishment period.

At the regional level, figure 42 shows that the higher prevalence of ICT4D can be observed in APR from IFAD5 to IFAD9, in NEN for IFAD4 and in ESA for IFAD8.

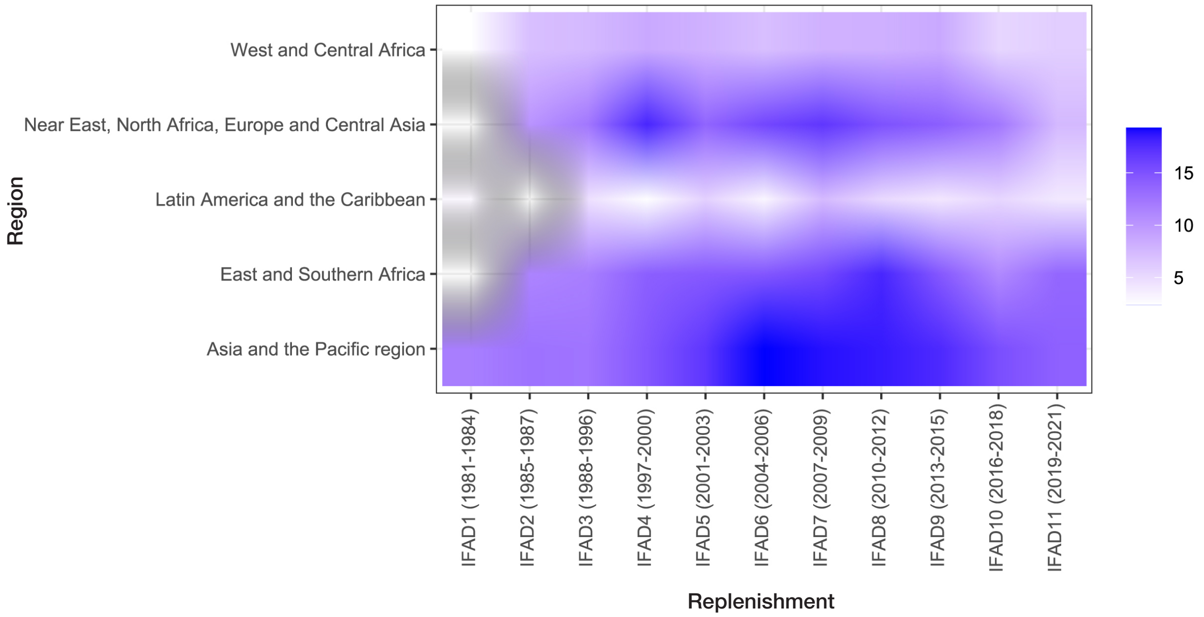


Figure 42 Presence of ICT4D taxonomy in IFAD project documentation, by region and replenishment period.

Overall, these results show that the presence of ICT4D-related terminology has been increasing in IFAD's documentation from 1981 to 2019. Regions leading this increase are APR and NEN. Lower presence of ICT4D-related terms has been observed in projects implemented in WCA and LAC. Figure 43 shows a break down by the focus area of the ICT4D solution, where a high presence of digital monitoring tools and digital advisory services are present in IFAD documents. Conversely, digital solutions focused on access to information are less present in the project documentation.

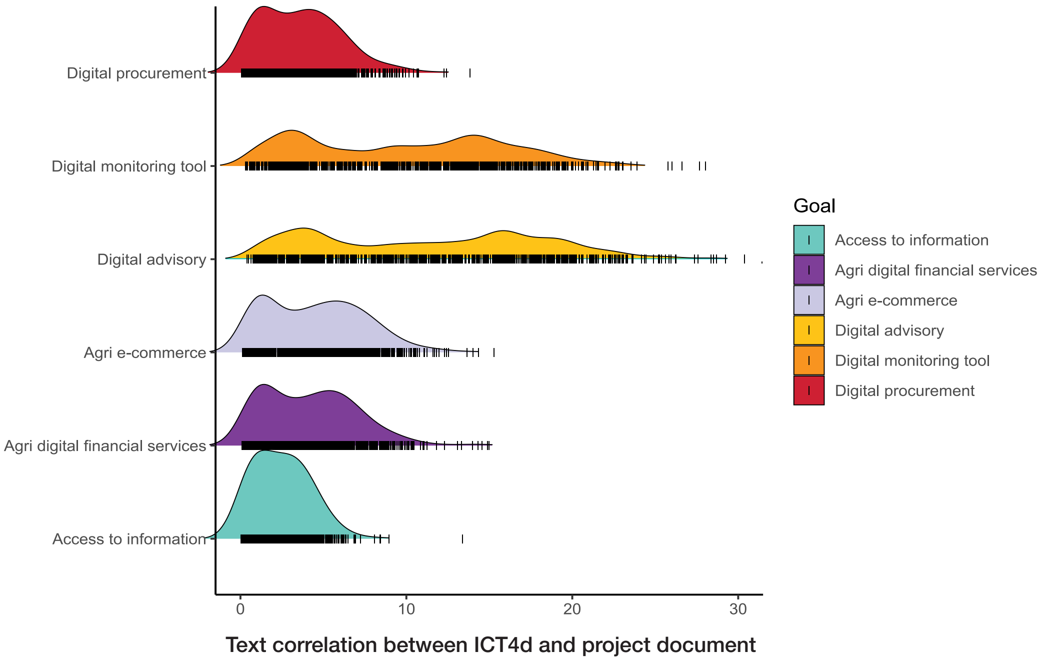


Figure 43 Text correlation between ICT4D focus area and project documentation.

Figure 44 shows the text correlation between the focus areas of ICT4D solutions and IFAD's replenishment periods. It confirms the prevalence of digital monitoring tools and digital advisory services, as they show the highest values from IFAD4 to IFAD9.

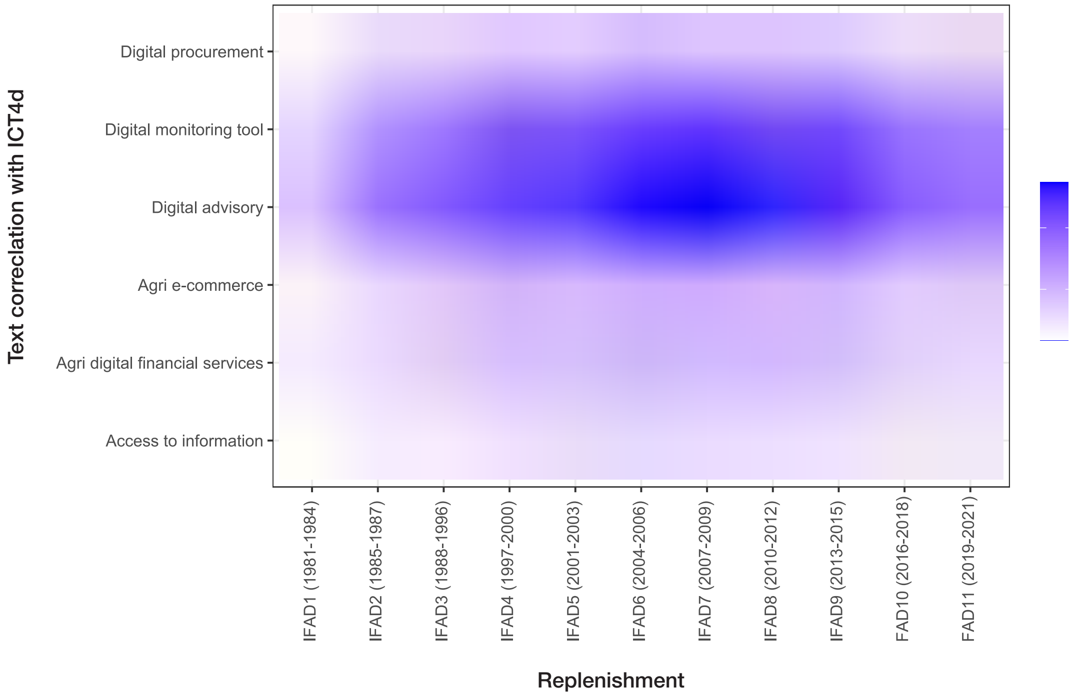


Figure 44 ICT4D presence in IFAD project documents, by focus area and replenishment period.

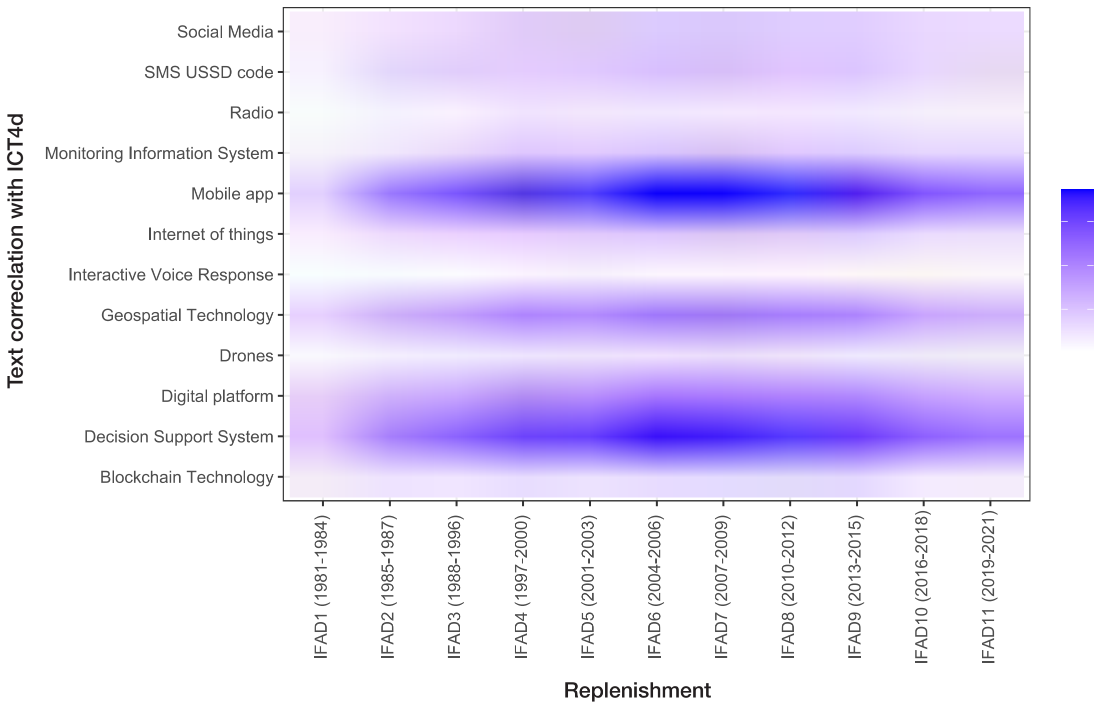


Figure 45 ICT4D presence IFAD project documents, by technology type and replenishment period.

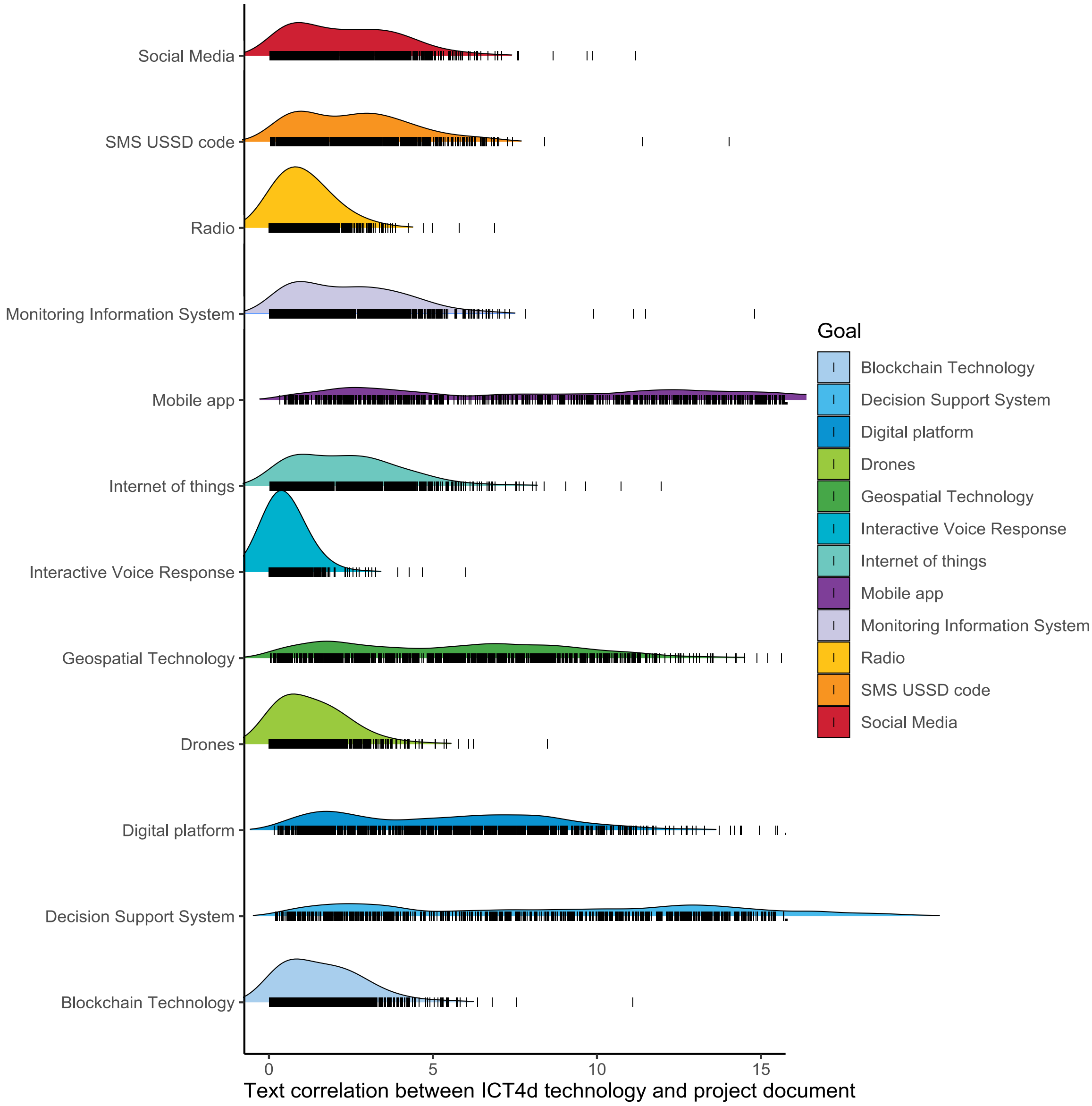


Figure 46 ICT4D presence IFAD project documents, by technology type.

The APR ICT4D solution draft proposal also identifies a typology for technologies implemented across projects. Their presence in the documentation is shown in figure 46 above. The technologies most detected are mobile applications, geospatial technologies, digital platforms and decision support systems. Radio, interactive voice response, drones and blockchain technologies were the least present.

Lastly, the technology types were also analysed against IFAD's replenishment periods. Figure 45 shows a peak for mobile applications and decision support systems in IFAD6 and IFAD7, that is, starting in 2004, which also reflects the internet's transition into greater connectivity and access through the Web 2.0.

Predictive analytics

In this work stream, which represents the core of machine learning applications, the project developed algorithms aimed at supporting the project cycle through ex-ante predictions of performance and probability of positive impact of IFAD-supported interventions, given a specific set of portfolio and beneficiary features. Two main prediction models have been built, at project and household level, respectively. While the first prediction can inform about successful features for portfolio performance and guide the organization on which projects are likely to succeed, the second can strengthen household level targeting at project design by determining the beneficiary and project level features that drive positive impact. We also responded to the COVID-19 crisis by proposing a ML approach to enhance knowledge about the impact of the pandemic in IFAD's beneficiary countries.

Impact of COVID-19 in IFAD beneficiary countries

COVID-19 (the novel severe acute respiratory syndrome caused by the coronavirus 2 "SARS-CoV-2") was recognized as a pandemic by the World Health Organization (WHO) on 11 March 2020, having caused economic and public health disruptions around the world. Addressing the pandemic requires economic and public health coordination at international, national and local levels, but the lack of reliable statistics is one of the most evident barriers to target the best policies at present.[9] Where official statistics are not readily available or reliable, the use of "big data" can improve our ability to understand and predict the evolution of complex phenomena.

This workstream aimed at estimating the real COVID-19 incidence in selected countries using big data and machine learning approaches, in order to support IFAD's understanding of the impact of the pandemic on beneficiary countries. A model was devised to predict regional spreading of COVID-19 in countries where IFAD operates and where official data is not available/reliable.

One of the most relevant information sources during the pandemic has been the database by the Johns Hopkins Coronavirus Resource Centre,[10] which is constantly updated with data concerning cases and deaths. However, the data has some limitations, especially related to under-reporting of new cases. Figure 47 shows how the distribution of cases per capita is uneven among across countries. Similarly, the deaths per capita are distributed in figure 48. In both instances, it is notable that in many developing nations there is a low and abnormal prevalence of the pandemic, indicating the possibility of under-reporting.

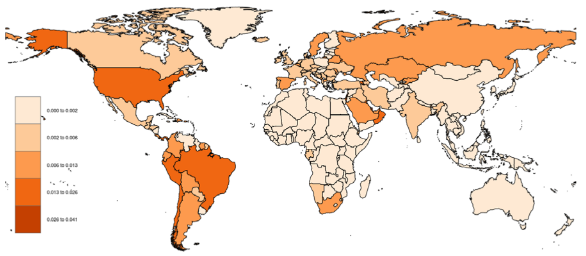


Figure 47 Cases per capita, as reported in the Johns Hopkins database (updated to September 2, 2020).

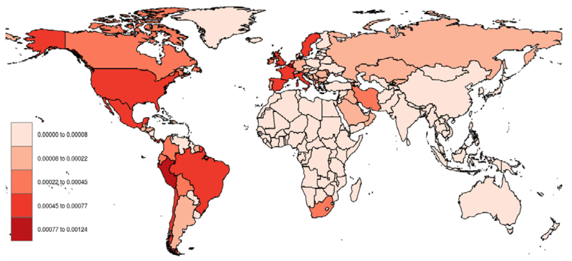


Figure 48 Deaths per capita, as reported in the Johns Hopkins database (updated to September 2, 2020).

Such a lack of data requires the development of solutions that correct the figures provided by Johns Hopkins. However, the use of data from internet search engines has shown high potential in the study of various cases of the spread of infections, with a significant impact on the "real time monitoring" of epidemics.[11] The idea underlying the approach

9 Brunori, P. & Resce, G. (2020) *Searching for the peak Google Trends and the Covid-19 outbreak in Italy;* Fantazzini, D. (2020) *Short-term forecasting of the COVID-19 pandemic using Google Trends data: Evidence from 158 countries.* Applied Econometrics.
10 https://coronavirus.jhu.edu/map.html

of Ginsberg et al. (2008) is simple: Internet users who suspect they have a disease tend to look for information online about symptoms and conditions associated with it. Such research leaves a trace, feeding the availability of a granular and massive dataset. A strong correlation between online search queries – provided by Google Trends – and influenza infections patterns has been confirmed by various studies and has now been applied to COVID-19. The maps shown in figures 50 to 54 present the distribution of symptoms related to the coronavirus, as represented by topics identified in Google Trends, namely: "coronavirus", "cough", "fever", "sore throat" and "pneumonia". Their distribution is clearly distinct from official data on contagions and deaths, with increased presence of the topics in African countries, which illustrates the hypothesis of a possible under-reporting.

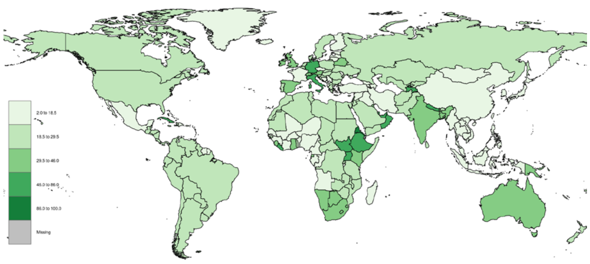


Figure 49 Distribution of queries related to the topic "Coronavirus" on Google Trends.

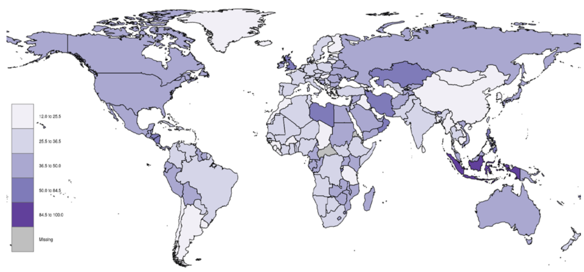


Figure 50 Distribution of queries related to the topic "Cough" on Google Trends.

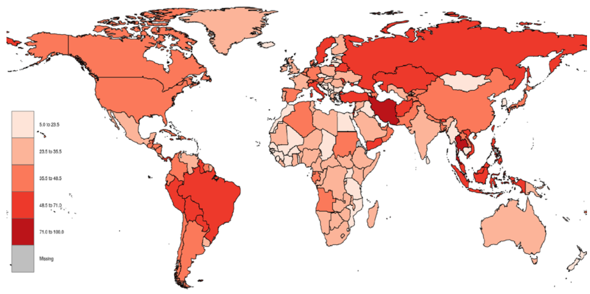


Figure 51 Distribution of queries related to the topic "Fever" on Google Trends.

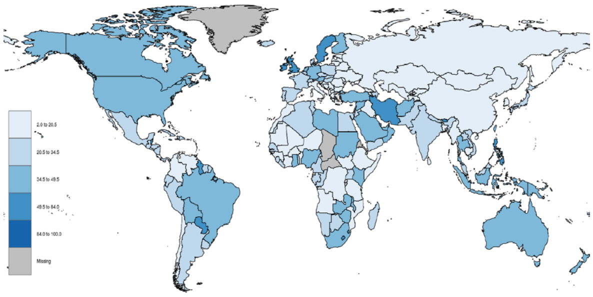


Figure 52 Distribution of queries related to the topic "Sore Throat" on Google Trends.

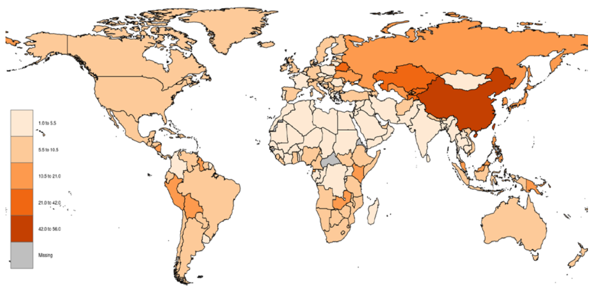


Figure 53 Distribution of queries related to the topic "Pneumonia" on Google Trends.

11 Ginsberg J. et al (2008) *Detecting influenza epidemics using search engine query data*. Nature 457: 1012-10155; Cook S. et al (2011) Assessing *Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1),* Pandemic.PLoS ONE 6(8); Broniatowski D.A. et al (2013). *National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic,* PLoS ONE 8(12): e83672.

In addition to Google Trends, another web source available from Google are the COVID-19 Community Mobility Reports. Google Maps uses aggregated and anonymous data to show how crowded certain places are, which allows identification of, for example, the peak hours of a restaurant. This type of aggregated and anonymized data could be useful for making critical decisions in the mitigation of COVID-19. Indeed, responses to the pandemic worldwide have increasingly moved towards public health strategies related to restrictions on movement and social distancing, in order to slow down transmission or plan re-openings. Information on mobility from Google is available for the following location types: food and pharmacies, parks, public transport stations, retail and leisure, residential areas, and workplaces.

To correct Johns Hopkins' official estimates with alternative data provided by Google, it is necessary to estimate a model that – as certain covariates vary – modifies the value of the COVID-19 cases and deaths. Given that most of the countries where IFAD operates are possibly under-reporting coronavirus cases and deaths, the objective is to predict the number of cases and deaths as accurately as possible, through the application of the "Random Forest" machine learning algorithm.

Johns Hopkins' data on number of cases per million and number of deaths per million have been analysed in relation to Google Trends topics, Google mobility reports, socioeconomic indicators, health indicators and governance indicators. After predicting for countries across the globe, predictions were compared with the observed cases, to highlight where there may be more deviation (i.e. under-reporting). The maps in figures 54 and 55 show that the most significant deviations refer to countries in Africa and Asia.

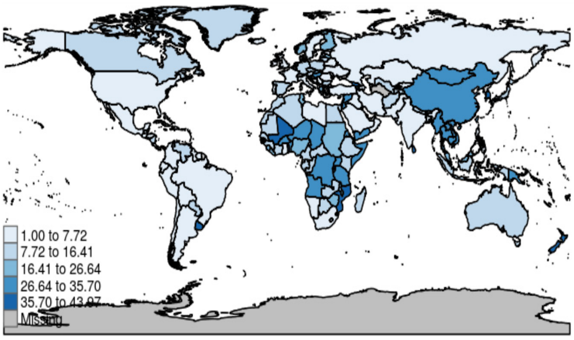


Figure 54 Intensity of predicted cases compared to official Johns Hopkins cases.

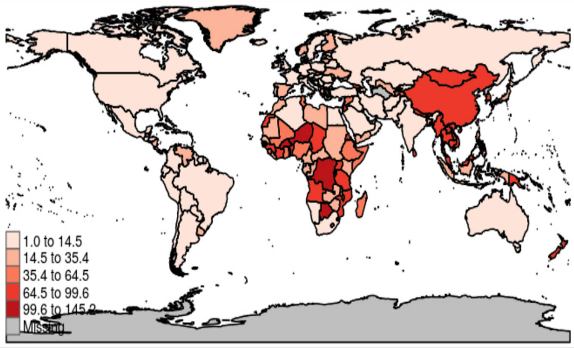


Figure 55 Intensity of predicted deaths compared to official John Hopkins deaths.

Countries that are identified as under-reporting in cases also tend to be those that are under-reporting in deaths. In order to test this association, a Pearson correlation was used, which is significant (p-value < 0.001) and equal to 0.75. As explained above, with Random Forest it is possible to identify which variables help to better predict the dependent variables. Regarding the number of COVID-19 cases, the most important variable is the number of tests performed, followed by the GDP per capita and the population. From Google Trends the search topics of pneumonia and fever are the most important, and as far as Google Mobility data is concerned, the best variable is the one referring to moves to the pharmacy.

In the case of deaths from COVID-19, the variables that best explain the number of deaths per million are health care spending and political management of violence. It is reasonable to think that lower expenditures in health care can mean lower capacity to treat COVID-19 patients, which increases the risk of deaths. Also, in this case, pneumonia and fever are the Google Trends topics that best help predict the number of deaths. As far as Google Mobility data is concerned, the variable that best explains the regression is park related mobility.

Using COVID-19 incidence predictions for planning

The output from this study is a tool that can help international organisations target countries needing additional funding due to the COVID-19 outbreak. It comprises a composite index that combines food security indicators with COVID-19 cases and deaths. The food security indicators considered are SOFI 2020[12] country level data for "prevalence of undernourishment in the total population" and "prevalence of stunting in children under 5 years of age". For cases and deaths, the predictions developed above were employed.

Composite indices developed through data-driven methods have been extensively employed as a technique for aggregation.[13] In particular, the data envelopment analysis (DEA) method compiles multi-dimensional metrics into one index using the combination of weights that is the most convenient for the evaluated alternative.

Using DEA, the global score for each country was estimated by a linear programme. Weights estimation can be difficult and highly subjective. In line with DEA methodology, the linear programme is computed separately for each country. The weights in the objective function are chosen optimally with the purpose of maximizing the score of the evaluated country. The optimization ensures that each country is evaluated on the basis of its own best possible weights. Figure 56 shows the association between the composite index estimated using predictions and the composite index estimated on actual COVID-19 data. Overall, the correlation is positive, although there are countries showing significant differences.

Focusing on countries where IFAD had projects, the countries with higher differences between the composite index combining food security measures and COVID-19 cases and deaths, before and after correction, are: Oman (+50%), Armenia (+18%), Montenegro (+18%), Cabo Verde (+13%), Suriname (+12%), Dominican Republic (+11%), El Salvador (+11%), Republic of Moldova (+11%), Macedonia (+11%) and Honduras (+11%).

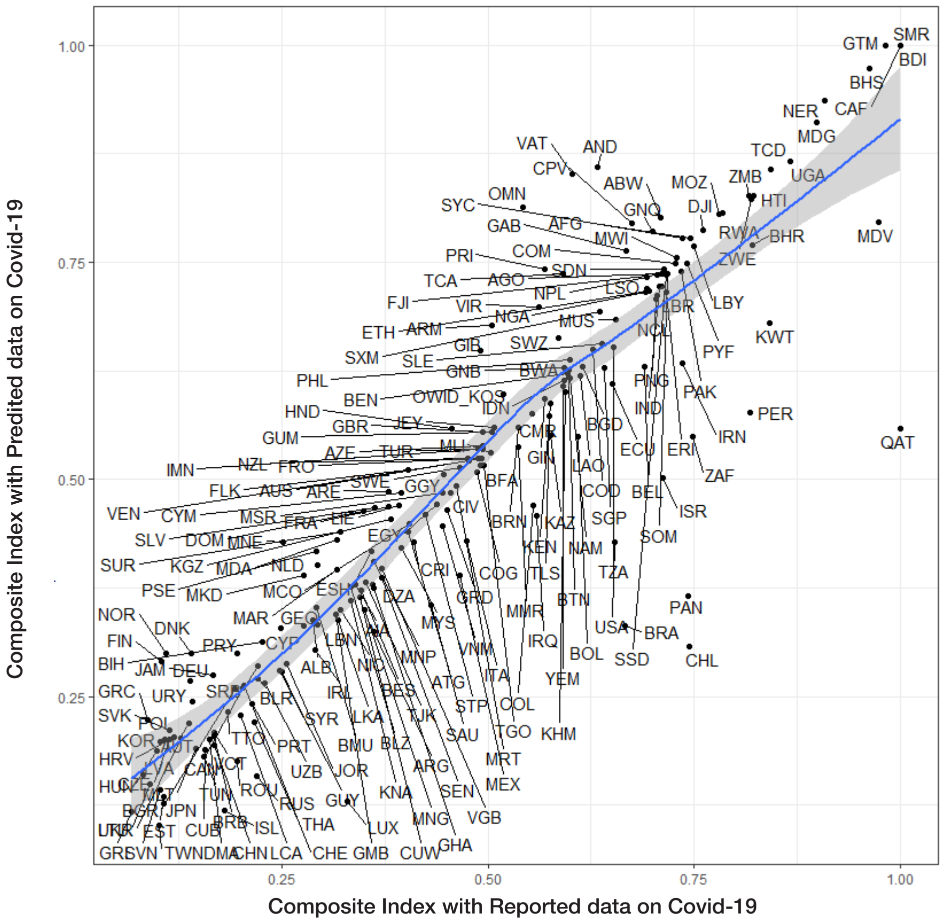


Figure 56 Correlation between the composite index estimated using predictions and the composite index estimated on actual COVID-19 data.

12 FAO, IFAD, UNICEF, WFP and WHO. 2020. The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets. Rome, FAO.

13 Decancq, K. and Lugo, M. A. (2013) *Weights in multidimensional indices of wellbeing: An Overview,* Econometric Reviews, 32(1):7–34; Patrizii, V. et al (2017). *The Cost of Well-Being.* Social Indicators Research, 133(3):985–1010; Greco, S. et al (2018). *On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness.* Social Indicators Research: 1-34.

Predicting performance at the project level

In the first phase of the project, a framework was developed to predict project performance proxied by disbursement readiness and other performance variables, such as ratings. Given that disbursements are linked to operational performance issues but currently also to financial capacity, the analysis focused on other outcomes (performance ratings only).

The prediction of performance is proxied by the rating at different points in time – at entry, during implementation and at completion – and identifies the drivers of performance at these different points in time for the entire portfolio. The analyses bring in macro data on poverty trends, food insecurity, share of GPD in agricultural sector, GDP growth and other covariates, that predict performance over time.

We present preliminary results on two main outcomes: (i) overall implementation performance; and (ii) likelihood of achieving development objectives. Results show that being either a potential or actual problem project, or being in country with a fragile situation, significantly reduces the rating values with a magnitude of around 0.90 and 0.20, respectively. On the contrary, increasing the share of projects financing to the livestock sector as well as total financing towards macro areas such as Social services, or access to markets, and the share of IFAD financing to areas such as access to markets and in macro areas such as environment, natural resources and climate significantly raises the rating values. According to these preliminary results, inclusive rural finance is the most financed macro area and has a significant positive impact, such that the more the macro area is financed the better the project rating – increasing the ratings value by about 0.43. Results are being refined and available upon request.

Predicting impact at the household level

Project impact assessment (IA) data was leveraged for the development of a prediction model that is able to detect the extent of positive impact based on project and household level features. Such an approach not only enables IFAD to gain further insight into its data, but also delivers added value as it can provide strategic information to policy makers concerning the effectiveness of policies.

Specifically, the model predicts the beneficiary features that optimize the impacts of specific policies. As such, it can help determine the types of interventions that predict the various positive outcomes the most, as well as the specific beneficiaries' profiles that are likely to exhibit more impact – given a specific policy or intervention. **This information is highly valuable for the design of new and effective projects and optimizes IFAD resources to increase its impact.**

Predicting Positive Impacts in IFAD10

Although primarily an ex-post evaluation tool, we still rely on prediction to estimate the share of beneficiaries receiving positive impacts, the likelihood of positive impacts, and the distribution of benefits among beneficiaries using household level data from a sample of IFAD10 households from 17 impact assessments.

For the policymaker, the tool is valuable because ex-post it allows for more detailed consideration of normative concerns, like equity, and identification of the features that contributed to more positive impacts. To demonstrate idiosyncratic effect estimation's utility for the policymaker, the team employed novel econometric techniques to estimate the conditional average treatment effect (CATE) function that maps beneficiary features to their programme impact.

The average treatment effect, which headlines most programme evaluations, provides an estimate of **programme impact** by calculating the difference between treated beneficiary outcomes and untreated beneficiary outcomes, controlling for selection effects and other confounding factors. Interest in this application is around the variation among individual beneficiaries as it facilitates estimation of the impact distribution and allows the team to expand the scope of our evaluation to consider whether the right people (however defined) are getting and benefiting from the programme.

For each of the 17 projects evaluated during IFAD10, the impact distribution on outcomes proxying for each strategic objective and the corporate goal were estimated. The results show that, even given imprecise mean impacts, there is substantial variation within and across projects in **how they distribute returns among their beneficiaries**. Moreover, it shows that comparison across indicators in terms of their distribution can provide more detailed insights than mean or subgroup mean comparison by allowing impact to vary across all levels of relevant covariates. In this case, focus was put on the share of beneficiaries receiving positive impacts: to demonstrate the utility of the model, the results on income were presented, the overarching goal of IFAD, and a within project comparison of results on a market access indicator, e.g. value of sales. Figure 57 shows the distribution by project for gross total income and gross agricultural income (proxy for corporate goal).

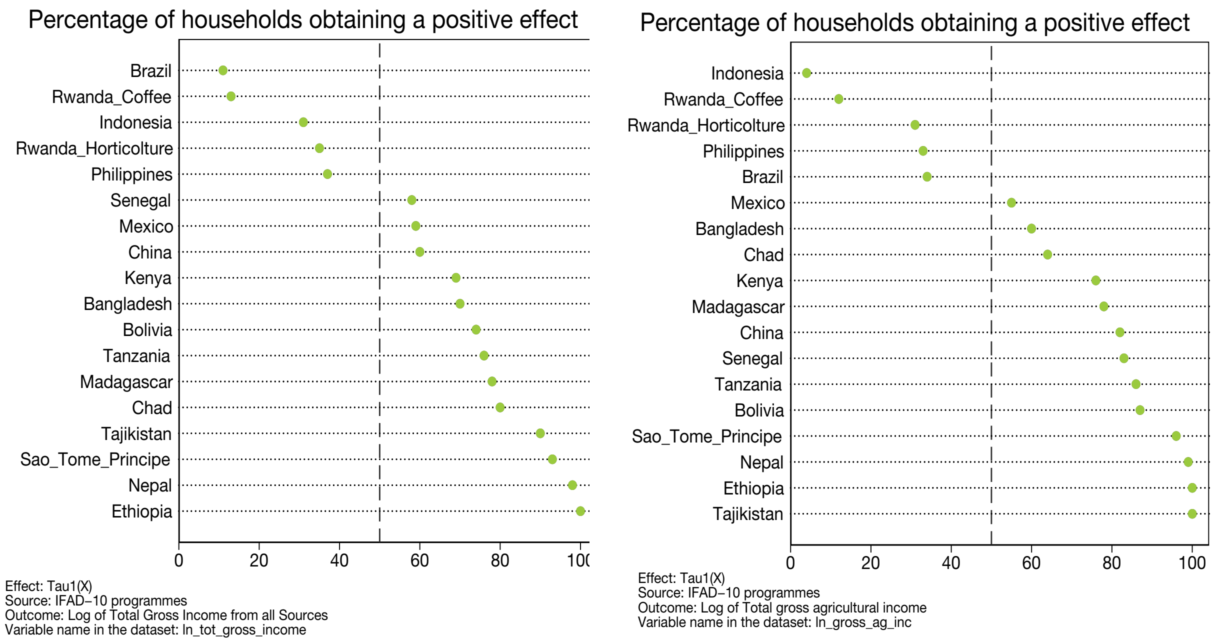


Figure 57 Share of beneficiaries with positive effects on total gross and total agricultural income

On income, IFAD10 projects were relatively effective in allocating benefits to a majority of their beneficiaries. All but three projects and the coffee component of Rwanda's Project for Rural Income Through Exports (PRICE) delivered positive impacts on total income to greater than 50% of beneficiaries; similarly, only four projects did not deliver positive impacts on agricultural income. For treated households across the 17 IAs, gross income increased on average by 11% compared to agricultural income which increased by 15%. The largest impacts on total income were in Nepal and Ethiopia and returned positive impacts to nearly 100% of their beneficiaries. On the other hand, the largest negative impact was in Brazil where only 12% of beneficiaries experienced positive returns on total income. Broadly, the most inclusive projects delivered the largest average impacts on income. However, comparing the distribution of effects on agricultural income against the distribution of total income shows that in some cases, namely in Tajikistan, the horticulture component in Rwanda, and in Brazil, increases in agricultural income did not guarantee increases in total income or vice versa.

In Tajikistan, only 87 percent of beneficiaries experienced positive impacts on total income despite all beneficiaries increasing their income from agriculture. This could point to a reallocation of resources away from other sources toward agriculture as a result of the project. Similar patterns were present in Brazil where increases in agricultural income did not convert consistently into impacts on total income. Conversely, for beneficiaries of the horticulture component in Rwanda, a larger share of beneficiaries saw positive impacts on total income compared to impacts on agricultural income despite the project's targeting of agriculture (figure 54).

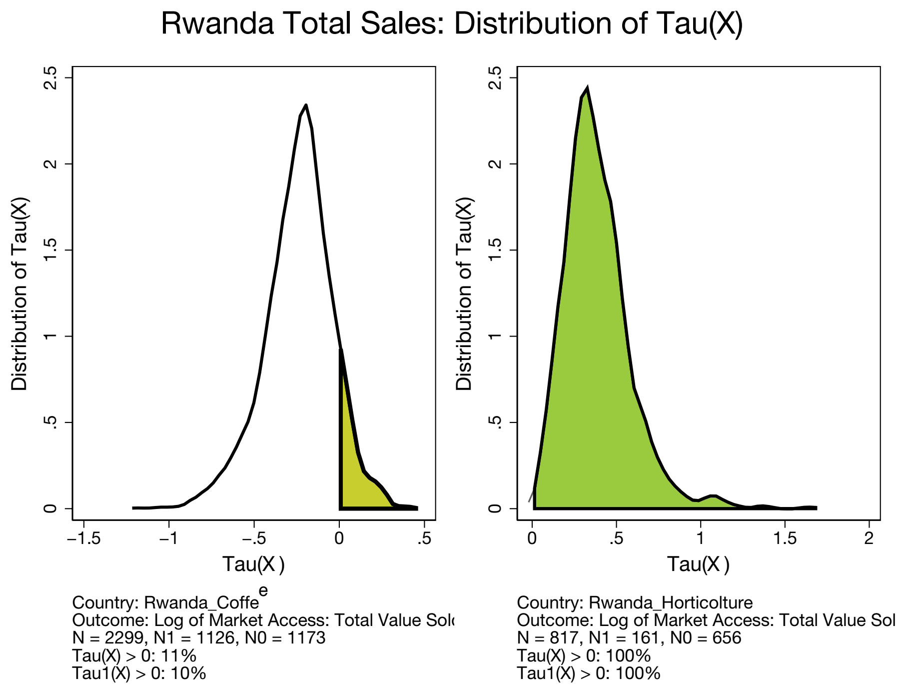


Figure 58 Rwanda Horticulture vs. Rwanda Coffee Distribution of Tau.

In addition, this work can be used to illuminate project and targeting effectiveness. Focusing on the two components of the PRICE intervention in Rwanda, both components aimed to increase sales for their beneficiaries but targeted very different populations. Horticulture farmers receiving matching grants were on average richer as the coffee component specifically treated struggling cooperatives for capacity building and input provision. On total sales, 11% of coffee farmers saw impacts compared to 88% of horticulture farmers. This difference is reflected in project average impacts, though considering only the average impact obscures how different the distributions are. For horticulture farmers, despite successfully delivering positive impacts there are a substantive number of beneficiaries for whom the project did not work as well. In this case, for policymakers looking to scale up or replicate the component, they would want to learn what factors drive this issue. For coffee farmers, results seem to suggest mis-targeting and misallocation toward ill-suited cooperatives or ineffective implementation. The original report from the IA notes systematic variation in implementation phases, but these results show that improvement in the second phase of implementation may have failed to offset negative effects in the first phase.

For policymakers, this application of ML allows for more nuanced breakdown of impact to identify what worked, what did not, and for whom. This is crucial for evaluation, especially when the policymaker has interest in scaling up or replicating a programme, which is often the case in IFAD.

Optimal Targeting

Another application of ML is optimal targeting. For example: an IFAD programme manager who is tasked with designing a new project that scales up activities conducted under a similar project in a previous period is faced with a targeting decision, namely who to select for participation in the new programme. The targeting decision can be restated as a constrained optimization problem: they must choose, given a set of budgetary, ethical and other policy constraints, whom to treat to achieve maximal aggregate impact and return. In other words, they seek a selection rule that maximizes the policies' total effect subject to existing constraints. The results of traditional IAs do not identify such a rule or solve the programme manager's decision problem. We link the work of the IFAD evaluator's assessments and the work of the IFAD policymaker through the application of optimal targeting to the Agriculture Sector Development Programme and Livestock and Agricultural Services Support Programme in the United Republic of Tanzania (ASSP/ASDP-L).

For the Tanzania application, focus is put on the case where there are two selection variables that will define programme participation, namely, age and education of the household head; and maximization of the impact on market access as measured by total agricultural sales is sought. The ASSP and ASDP-L programmes were designed and implemented to develop agricultural production systems through capacity building and farmer field schools (FFSs). FFSs used a participatory approach in which trainers facilitate farmers' learning and problem solving and promoted new techniques with the goal of increasing beneficiaries' productivity and their ability to access markets. Training topics included the use of artificial insemination, calf rearing, linkages between farms and markets, land preparation, use of manure, organic farming, promotion of highly nutritious crops and climate adaptation. Given that the path to impact hinges on beneficiaries' understanding and adoption of new techniques and practices, age (as a proxy for experience) and education are likely to condition impacts. And, as they are easily observed and collected ex-ante, they function well as selection variables.

The objective of the model is not to identify a single solution, but to provide the policymaker with a menu of options such that they can identify the trade-offs among different targeting rules. Some assumptions, however, are made: only rules that target treatment to at least 20% of the target population are considered; and feasibility constraints are placed on age and education thresholds, such that only age thresholds less than 65 years and education thresholds less than 16 years are considered. The policymaker can apply additional constraints. For example, the programme may endeavour to be youth-inclusive so the policymaker may only accept thresholds that are less than 20 years. In these scenarios, the model's utility is readily apparent because the trade-off the policymaker must make to be inclusive of households headed by young people is obvious. Additionally, budget constraints are easily applied and the trade-offs between costs and impact are apparent.

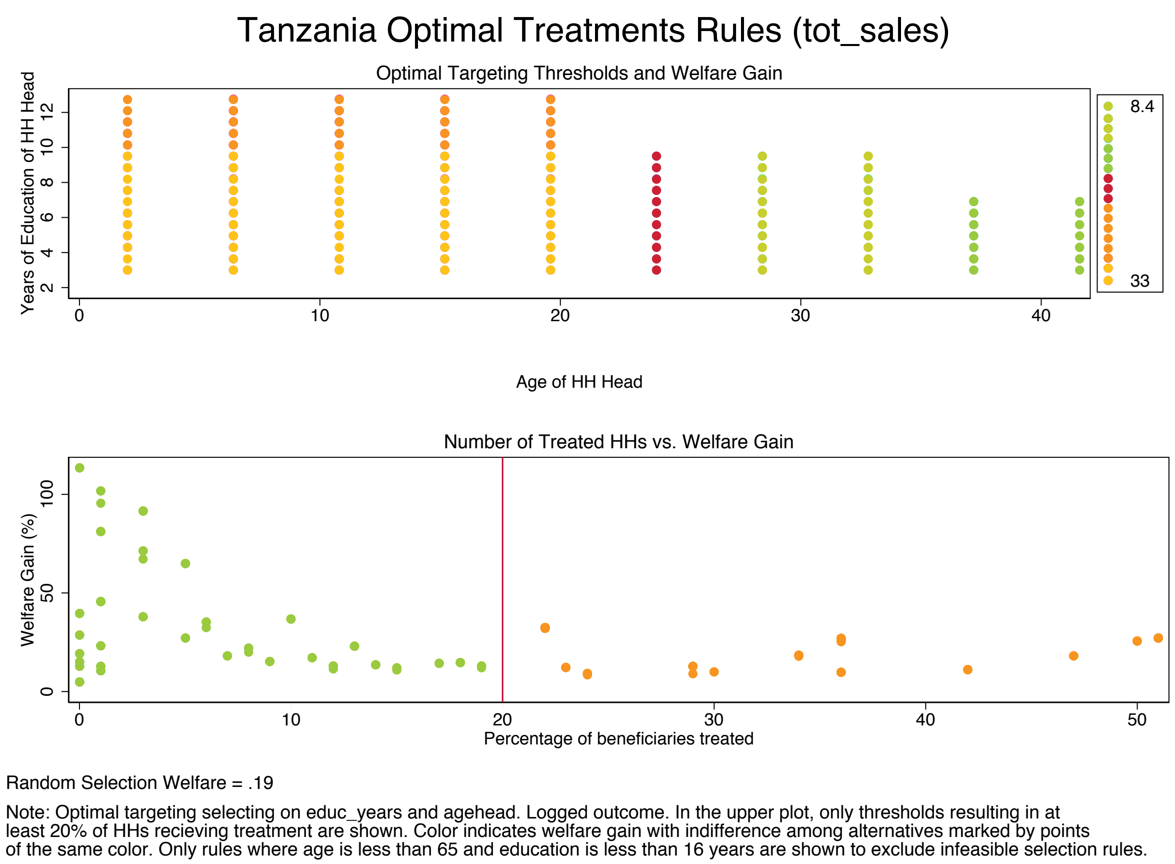


Figure 59 Optimal thresholds to maximize impact on total sales.

Figure 60 shows the results on total sales, the market access indicator from the model. The top plot shows the aggregate welfare gain (proxied by market access) from each treatment rule. As expected, as education increases the welfare gain increases but contrary to expectation as age (or experience) increases the welfare gain falls. When considering feasible thresholds, the maximum welfare gain is 32.5% increase over random selection with an age threshold of just 2 years and an education threshold of 10 years with 22% of the population treated. Of course, there are no households headed by two-year-olds, but this shows that impact is mostly gained through selection on education. There may be cases where being youth inclusive but selecting on education is an acceptable trade-off, however as educational attainment is correlated with wealth, location, and gender (among other factors) the policymaker may decide such a threshold is too discriminatory. As shown in figure 60, lowering the education threshold below 10 years necessitates lower welfare gains. To achieve a welfare gain of 27% while being inclusive of households headed by youth (age 18 to 25), the policymaker must also elect to treat a larger number of beneficiaries, fifty-one percent. In these cases, the model makes it readily apparent that in order to be more inclusive along both selection dimensions the policymaker must accept lower welfare gains and a higher number of beneficiaries.

The presentation of these results is not meant to suggest that age and education are the most relevant dimensions on which to target, nor do we recommend any specific trade-offs or treatment rules. The purpose is to illustrate that with relatively little information on prospective beneficiaries, and given relevant rigorous micro-level data, there will be a menu of options to policymakers such that they can make evidence-based decisions with significantly improved awareness of those decisions' costs and benefits.

Policy implications

From a corporate perspective, as IFAD's Strategic Framework 2016-2025 highlights the centrality of ICT and encourages expanding the uptake of new technologies, this wide set of use cases or innovations foster the creation of tools to enhance knowledge management and support IFAD's ICT4D strategy by proposing an **integrated, machine-driven approach to analyse project documentation, predict impact and inform the design of new operations.** It also addressed concerns from the IFAD Knowledge Management strategy regarding the fragmentation of information, by showing how new technologies help leverage existing data sources to answer new questions (for instance the different topics, interventions and lessons learned covered in the entire portfolio).

The Athena project has essentially aimed to repurpose existing data to gather new insights – notably both project documentation and micro-level data from impact assessments. Text mining and topic modelling can turn masses of unstructured text from IFAD documents into structured data, which can be then analysed for trends, integrated with other data sources, and incorporated into machine learning models.

In addition, household data from IAs have been repurposed for the development of a prediction model that is able to detect the extent of positive impact based on project and household level features. Such an approach not only enables IFAD to gain further insight into its data, but also delivers added value as it can provide strategic information to policy makers concerning the effectiveness of policies. Specifically, we can predict what are the beneficiaries' features that optimize the impacts of specific policies. For instance, ML can help us determine the types of interventions that predict the various positive outcomes the most, as well as the specific beneficiaries' profiles that are likely to exhibit more impact – given a specific policy or intervention. This information is highly valuable for the design of new and effective projects and optimizes IFAD resources to increase its impact.

The development of a framework that predicts project performance and identifies key drivers linked to project success, as well as positive or negative impacts of interventions, can guide policymakers about the effectiveness of planned policies and interventions among beneficiaries to help target interventions more effectively in future projects. Specifically, these methods allow for increasingly nuanced takeaways from IAs that can be applied in future projects. With the applications of these techniques, not only would we know, for example, that on average a project increased income, but we would also know through what means it increased income, such that future projects can maximize impacts or determine if impacts were distributed differentially among the beneficiaries. In the field of evaluation, the impact of ML is most acute for the evaluator in this regard, because ML allows for better ex-ante prediction of individual impacts such that decision rules can be applied effectively, conditional on these better predictions. These decision rules would lead new projects to be designed in ways that maximize impact while minimizing resources, as well as guide country strategies on the best opportunities for impactful and cost-effective interventions and targeting mechanisms.

Therefore, the applications for ML employed in the Athena project greatly support IFAD's Development Effectiveness Framework, especially regarding improving efficiency in corporate reporting. In addition, such tools contribute to building an evidence base to inform policy and the design of successful projects and country strategic opportunities programmes. In addition, they bring in principles of cost-effectiveness and value for money in the context of project design, ex-ante.

Additionally, Athena has indicated how ML can play a key role in supporting the advancement of IFAD11 commitments and targets by accelerating project-level data analysis, thus **enabling more regular reporting on the mainstreaming themes and the SDGs against the strategic outcomes.** By harnessing internal and external evidence of which interventions deliver the most effective results, not only can IFAD strengthen its own project design, but also disseminate learning to other international institutions. As shown above, a number of ML algorithms have been developed to identify themes and features within project documentation (food systems types, SDGs-related content, mainstreaming themes, and ICT4D features), with methods that are flexible enough to adapt to any strategic theme to IFAD. This approach supports policy and development effectiveness by enabling a global picture of how IFAD invests in a particular topic, based on information from project documentation.

In summary, achievements of the second phase of this project have several implications from a policy perspective. Machine learning approaches were tested to systematise/integrate different sources and type of data to accelerate knowledge generation for data-driven operations. These outputs can provide an added value in:

(1) **Aiding and simplifying IFAD reporting** (for instance the IFAD Report for Development Effectiveness – but also more complex and data hungry thematic reporting (for instance concerning mainstreaming themes, food systems components and extent of reporting against SDG targets and goals in IFAD projects).

(2) **Enabling ex-ante data driven design**: **such tools have the potential to inform the design of new operations particularly country programme strategies (COSOPs), and design documents** about cost-effectiveness of interventions proposed as well as impact and performance potential.

(3) **Informing ex-ante, targeting strategies** by proposing a menu of options for assigning projects interventions in order to have impactful and cost-effective operations.

(4) **Enhancing development effectiveness** by contributing data-driven and evidence-based analytical solutions at each step of the project cycle.

(5) **Supporting the ICT4D development strategy** which foresees artificial intelligence, machine learning and big data as key pillars. Integration of ML and AI activities within ICT4D action plan should be sought in the future.

(6) **Contribute to the knowledge management action plan**, by aiding the dissemination of knowledge to users in a user-friendly way, (e.g. for instance through the development of dashboard, web apps and automated briefs as well as through the portfolio systematization workstream).

Lessons learned

- **Sustainability and Open-Source AI/ML:** In order to really transform the business model, the sustainability aspect of the innovation is a key factor to ensure the innovations' uptake by the organization itself. The key metric for a successful innovation is in fact its adoption and internalization by the organization as a whole. In this context, sustainability means being able to sustain and scale up the use cases after the prototyping phase has concluded. For IFAD, this would mean being able to embed the use cases in the overall business model, so that they don't remain stand alone or static products. Open-Source AI/ML and code transparency are essential elements that ensure that dashboards and apps can be updated in a real time fashion, when new data comes in as well as integration with the organization's data ecosystem.
- **Human element to improve algorithmic performance:** the human element is essential to improve the accuracy of algorithmic performance and overall quality of the models. In the case of the AI-based intervention dashboard, IFAD staff and domain experts have provided accurate taxonomies and training datasets that have fed the models, producing classifications that are "realistic". Additionally, IFAD staff have validated the results manually to ensure that the model has learned "correctly". While machine-learning models open up the possibility of automating analysis of large quantities of data, human validation is a key factor to establish accuracy and helping calibrate the models. As such, buy-in from key users is a crucial for the success of any prototypes.
- **Users' validation of new technological tools requires time**. First, users need to learn, explore and feel comfortable with the tool, to then be able to validate the actual data presented. In the case of the AI-based interventions dashboard, while a detailed manual, a video and live sessions were held to present the dashboard to key users, further technical support and training would be beneficial to encourage more involvement.
- **The complexity of IFAD project documentation** is a key challenge for models that require standardized data. Not only are project reports written in four different languages, but they also vary in format and length (variations in templates over time). Data processing requires the development of multi-lingual algorithms and sensitive data filtration strategies to ensure relevant text is extracted for analysis.
- **Establishing a continuous partnership with ICT**: Sustainable and sustained AI/ML and "big data" use cases require appropriate data repositories, server space, and secure data storage within the business model.
- **Procurement related aspects**: Digital development and Ethical AI[14] principles need to be embedded in developers' contracts to prevent the innovation to be commercialized for profit.
- **Modus operandi or business model while developing innovative products**: programmers or developers, statisticians, economists and subject matter experts have to work together intensively while developing products. Machine learning has always only been a serving role in the processes in development. It can assist the subject matter experts and management in their decisions.
- **Patenting of innovation**: developing innovation prototypes in partnership with third parties requires patenting of innovation.
- **Data protection**: use of AI and ML techniques requires the need to refer to data protection and data governance policies. The latter need to be in place to prevent data misuse.

14 https://en.unesco.org/artificial-intelligence

Recommended next steps

Currently, there is no centralized location that can both allow users to source information to document the cost-effectiveness of interventions and provide granular information on IFAD project features across the whole portfolio, nor is there a system that returns evidentiary sources on the likelihood that a development intervention will generate impact or a positive return. As part of the next phase of Athena, the acceleration phase (Phase 3), the project team **proposes both the expansion of the information generated and the integration of the various AI tools developed into a one-stop shop platform**, which would be made available across IFAD. The platform would be instrumental to aid data driven design of new projects and COSOPs, but also to aid reporting and any other information needs. For a successful acceleration of this project, the project will be led by the Programme Management Department (PMD) to better tailor the use cases to specific operational demands and user needs, additionally closely partnering with the Information and Communications Technology Division (ICT) to seek support with data storage and web hosting and ensure compliance with data systems and solutions within the organization. Also, further consultation with users will be key for the validation of results and for continued testing of the prototypes, also for a relevance perspective.

The tools produced need to be embedded within current ICT systems so that periodic updates can be sourced directly from existing reporting mechanisms. Specifically, a next step for mainstreaming the AI-interventions dashboard and the Lessons Learned web app would be linking algorithmic procedures and integrating them to IFAD Results Management systems, such as ORMS/Oracle BI maintained by the Operational Policy and Results Division (OPR).

Facilitating this data integration would make it possible for users to quickly manipulate and access the content and query the database for their needs while being connected to IFAD systems, which are updated on a regular basis.

For this project's acceleration, efforts should be concentrated towards both the systematization of the portfolio workstream, e.g. to improve, automate and streamline IFAD reporting, and the predictive analytics workstream, which is particularly useful in an operational context, and specifically for ex-ante project and COSOP design. Predictive analytics entails having a framework that allows one to have different prediction models for different "objectives" (e.g. project performance, impact and targeting), producing ex-ante analyses where the value added can also incorporate a cost-effective element.

Based on the results of the Phase 2 of the Athena project, the acceleration phase will foresee consolidating and scaling up the application of ML in IFAD's operations in the following ways:

1. **Integrate the use of machine learning within IFAD systems**, to be able to improve operational effectiveness by responding, in a timely manner, to the different information needs of the organization and its users; this may include the following activities:
   (a) **Capitalize and expand on the information generated** to ultimately condense in to a one-stop-shop the information gathered across the entire portfolio on the following topics: mainstreaming themes, SDGs and targets, food system dynamics, and extend the possible modelling to other themes, among others, topics related to persons with disabilities.
   (b) **Expand and further refine the Lessons Learned App,** a sustainable open-source solution. The next step would be to test and validate it with the Knowledge Management group and other users, specifically the Operational Policy and Results Division (OPR), to ensure relevance and adoption of the tool. Additionally, more documentation which contains lessons learned from IFAD operations, should be integrated, to create a taxonomy of lessons learned that can inform feedback processes across operations.

(c) **Leverage the information generated so far, to produce automated briefs** which could inform project design and supervision missions; the information for the briefs could be sourced from either the intervention dashboard and other databases available in the organization. Cognizant of the fact that users may need tailored and possibly differentiated information, it is worth engaging further with selected users to fine-tune the prototype and understand demand for specific products.

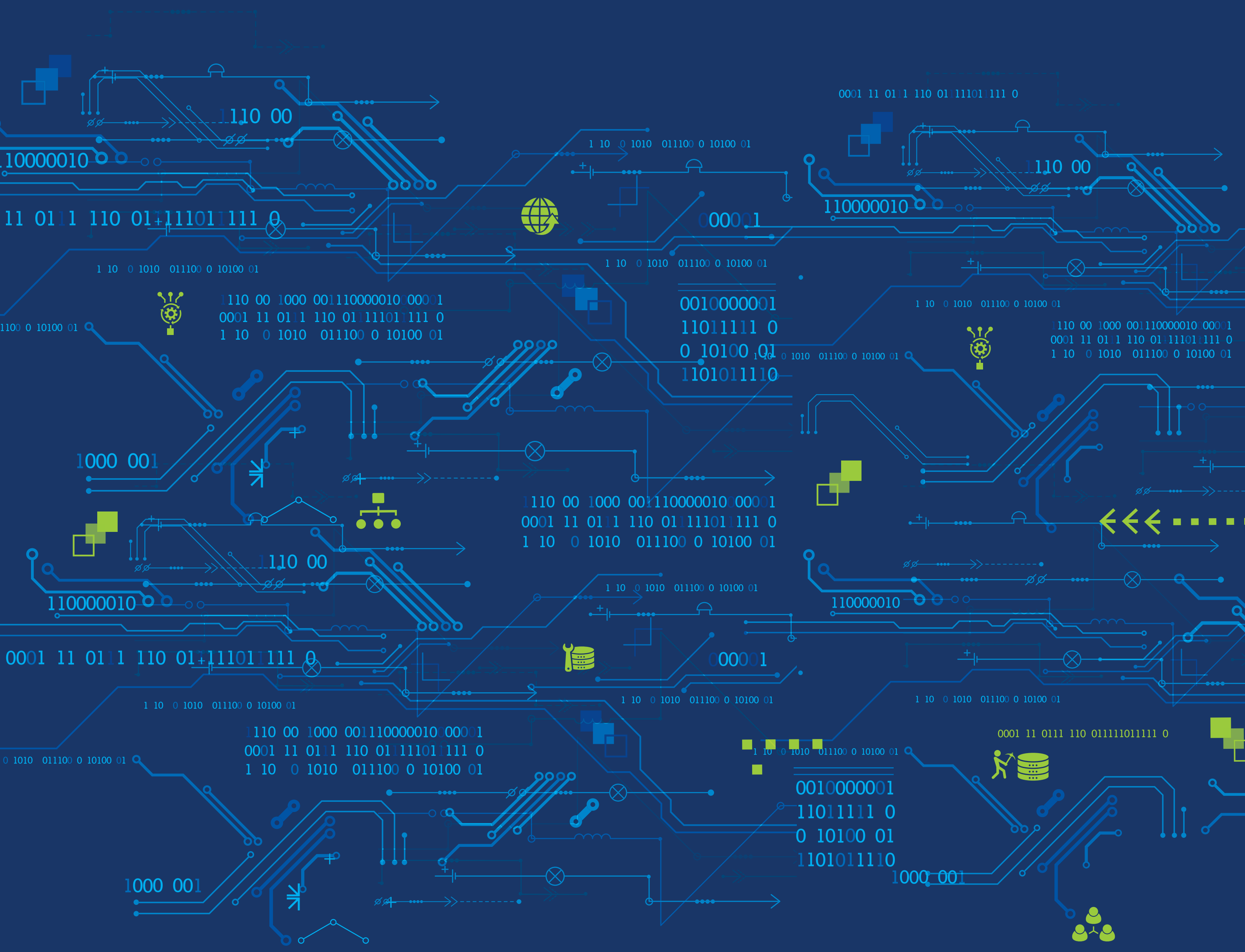
2. **Use of ML for impact evaluation**: the Covid-19 impact prediction model will be used in the context of assessing the impact of the pandemic in IFAD beneficiary countries on macro economic indicators (poverty proxies, GDP trends, food security etc). The scope of this tool would be to also prioritize financing needs in most affected countries.

3. **Using ML and big data to address targeting needs and to better target development efforts of new operations**:

   (a) Unconventional datasets, including mobile phone transactions, and remote sensed data can be integrated with secondary data and more broadly censuses and household surveys to update poverty and food insecurity maps and therefore aid targeting in countries where data are insufficient. This is the direction that other development partners have taken in a number of countries to better target development efforts.[15]

   (b) Additionally, ML can be employed for geographical targeting of infrastructural projects based on remote sensed data: IFAD can capitalize on the efforts of the ICT division, which created a systematic repository for GIS data (Geonode). The latter can be used in conjunction with ML for targeting infrastructure needs,[16], for instance.

15 See https://blogs.iadb.org/ideas-matter/en/when-big-data-and-machine-learning-spur-development-efforts/).
16 This is also the direction of the IDB where a “Building Detection Model” was created to generate a basic map of buildings from satellite images. This aided urban planners to create detailed digital files of hard-to-reach and remote areas, such as the Guyana hinterland, where their development were planning for housing interventions (see https://blogs.iadb.org/ciudades-sostenibles/en/urban-machine-learning-automatic-classification-of-buildings-and-structures/).



IFAD

International Fund for Agricultural Development
Via Paolo di Dono, 44 - 00142 Rome, Italy
Tel: +39 06 54591 - Fax: +39 06 5043463
Email: ifad@ifad.org
www.ifad.org

facebook.com/ifad
instagram.com/ifadnews
linkedin.com/company/ifad
twitter.com/ifad
youtube.com/user/ifadTV