



DIGITAL FINANCIAL SERVICE FOR SMALL HOLDER FARMERS

What data can financial institutions bank on?

November 2017



**MERCY
CORPS**

**AGRIFIN
ACCELERATE**

AgriFin Accelerate is a six-year initiative supported by the **Mastercard Foundation**, working in Kenya, Tanzania and Zambia.

TABLE OF CONTENTS

Introduction	3
Building a Digital Data Ecosystem	3
Successful Digital Loan Products Require More than Data	3
Where Does the Data Come From? 5	4
What Types of Digital Data Can Financial Institutions Bank On?	4
Credit Scoring	4
How Credit Scoring Models Work	4
What Kinds of Data Speak to the Risks of Small-Holder Farmers?	9
Purchase/sales records	9
E-learning on relevant topics	10
Farm-Related Geospatial Data	11
The Potential Value of Self-Report Ag-Related Data	12
Potential Use of Other Non-Ag Related "Alternative" Data	12
Psychometric Testing	12
Social Media Data	13
Putting it All Together	13
Data Collection Strategies to Facilitate Scorecard Development	13
Working with Commercial Partners	14
Consent, Data Privacy and Fraud Protection	14
Business Requirements/Service Level Agreement Terms	15
Conclusion	15



Introduction

Digital services create data. The past few years have seen an increase in innovation in digital products and services aimed at small holder farmers. Mobile devices and operator networks create 'digital delivery channels' to bridge geographic distances and bypass other obstacles (poor roads and infrastructure) to deliver financial, informational, and market facilitation services to the millions of rural farmers in Sub-Saharan Africa.

As more digital services are made available to small holder farmers, data collected on their characteristics and activities creates opportunities for financial institutions. The analysis of purchase and sales behavior, location-based information, and farm and farmer characteristics can inform more accurate measurement of the risks and rewards of lending to farmers, a population that until now has been 'bankable' only through their own cooperatives or hands-on, labor-intensive microfinance institutions.

Building a Digital Data Ecosystem

The first step in enabling digital data credit scoring for farmers is in creating relevant data sets on farm characteristics and behaviors. The more farmer-level data that is consistently collected, the better lenders will be able to assess and price the risks of lending to them.

The AgriFin Accelerate (AFA) network ecosystem includes mobile network operators, financial institutions, farmer networks, technology innovators, agriculture value chain players, government and other key market stakeholders. The digital services these actors provide include digital training on farming practices to transportation logistics, input purchases, market pricing and market linkages for the commodities produced. Each of these activities creates a set of potentially relevant digital data. The challenge is to find ways for ecosystem actors, whether individually, collectively, or in partnership with third parties (fintech companies, financial institutions), to make sense of the data and what it says about a farmer's likely credit risk.

Successful Digital Loan Products Require More than Data

Highly successful (indeed 'transformational') digital savings and loans products, such as M-Shwari in Kenya, have given cause for optimism about the possibility of using digital data to lend to small holder farmers. Where there is a will (and risk appetite), there will be a way, yet it remains true that risks and cash flows patterns in agriculture and livestock production are much tougher to algorithmically measure and match with financing than an individual's short-term consumption needs (very small loans of up to 30 days are the common 'nano-loan' product structure). This means that, in addition to data-driven credit scoring models, which are the focus of this paper, first movers in digital financial services for farmers will need farming knowledge sufficient to design products and processes that:

- Match loans' cashflows to the harvest and/or livestock business cycle
- Mitigate risks of loan default due to crop failure and/or livestock disease
- Are profitable and sustainable for the financial institutions
- Look out for the interests of local communities, perhaps through cooperation with insurers, cooperatives or other local support networks.

This paper looks next at where digital farm data comes from. It then considers how traditional credit scoring methods can use this data to measure the risks and rewards of lending to small holder farmers. Finally, it looks at other areas of sensitivity and risk around digital farmer data and credit, including farmer consent and data privacy, fraud, and the roles and responsibilities of business partners (such as software and scoring vendors).



Where Does the Data Come From?

Data is created each time a farmer engages with a digital platform. Some simple examples include:

- Date and time of interaction with the digital platform
- Record of the service provided (information accessed, purchase/sale made, training course accessed, etc.)
- GPS location of the mobile device accessing the service.

Digital service providers should capture, organize and store this data in ways that facilitate its future analysis, as data analytics are key to better understanding and serving their customers.

One promising use of digital data is simply to put small holder farmers on the radar of financial institutions. Third-party verifiable information on farmer identity, farming track record, and evidence of cash flows (of input purchase, crop sales) give financial institutions a potentially affordable and reliable way to assess a farmer's credit-worthiness via digital channels. Financial institutions are advised to start small, with targeted and tightly structured standardized products that should limit risk and help farmers build credit histories in order to unlock larger or more complex loans in the future.

What Types of Digital Data Can Financial Institutions Bank On?

Even amid the constant media and business-press excitement about "Big Data", "Machine Learning" and "Artificial Intelligence", the task for financial institutions assessing credit worthiness has remained rather straight-forward—to find evidence that the potential borrower is both able (has sources of income) and willing (has honored past obligations) to repay the loan.

There are many types of data collected over digital platforms that can help financial institutions understand a loan applicant's ability and willingness to repay a loan. Three of main types likely to be related to credit risk are:

1. **Verification of cash-flow:** such as records of sales and purchases over digital platforms; use of mobile money; use of air-time.
2. **Track-record (stability) in business:** length of time using a service platform, transacting with other members of the value chain, living in one location etc.
3. **Honoring past obligations:** payment of past credit obligations, bills, etc.

The next section looks briefly at how credit scoring works to understand what types of data can be useful in scoring models.

Credit Scoring

Credit scoring models are used to make-decisions for retail/mass-market credit products. In digital financial services, credit scoring models are used to make lending decisions on relatively small loan to borrowers the financial institution may have never before met.

An often-cited successful use of credit scoring in digital finance is CBA Bank's M-Shwari savings and loan product (in Kenya). M-Shwari clients receive instant decisions over their feature phones based on a credit scoring model that evaluates their past airtime and mobile money usage. This and competing similar products have significantly increased the share of the population with a bank account and/or loan in the formal financial sector.

How Credit Scoring Models Work

Credit scoring models are based on the relations between the characteristics and behaviors of past borrowers and their loan repayment. They assume that new applicants who look like past applicants will also repay like those past applicants (on average). In other words, they assume the future will be like the past.

To illustrate how scoring works in the context of digital financial services, consider a mobile network operator (MNO) that offers "air-time credit", or air-time loans, to its clients who normally pre-pay for services. Air-time credit means that when the subscriber runs out of money, the MNO immediately provides some additional air-time, and the client must pay for it within 10 days. If the air-time is repaid within 10 days, the MNO considers it a 'good' credit. Otherwise, it is a 'bad' or delinquent credit (i.e. repaid after 11 or more days).

To build a scoring model, the MNO gathers data on 10,000 air-time credits (further we will call these simply 'loans'). The data set includes the subscriber's:

- Client ID Number
- Gender
- Date of Birth
- Date of Joining the Network
- Number of days to repay the air-time loan

Each of the 10,000 loans are coded 'good' or 'bad' based on the number of days to repay the loan—up to 10 days are 'good' and 11 or more days are 'bad'. Table 1 presents how the first 10 rows of the resulting data set might look.

Table 1: Data Set for MNO Credit Scoring Model

Row	Client Id	Gender	D.O.B	Reg Date	Days To Repay	Status
1	1003	M	8/8/1964	2/9/2017	3	GOOD
2	1213	M	9/11/1986	12/21/2011	8	GOOD
3	1543	F	6/16/1973	1/21/2013	2	GOOD
4	1793	F	3/8/1985	5/21/2015	12	BAD
5	2063	F	1/27/1980	9/9/2016	10	GOOD
6	2333	M	4/10/1967	2/21/2014	2	GOOD
7	2603	F	12/10/1948	2/21/2015	3	GOOD
8	2873	M	1/27/1980	7/21/2016	7	GOOD
9	3143	F	9/27/1978	5/9/2015	8	GOOD
10	3413	M	10/4/1965	5/9/2016	2	GOOD

To build scorecards, summary tables call cross-tabulations compare the counts of good and bad loans to another selected column from the data set. For example, Table 2 is a cross-tabulation of Gender and Loan Status.

Table 2: Cross Tabulation of Loan Status and Gender

	A	B	C	D (=B/C)
Gender	Goods	Bads	Total	Bad Rate
Male	3,877	657	4,534	14.5%
Female	5,123	343	5,466	4.1%
TOTAL	9,000	1,000	10,000	10.0%

To illustrate how scoring works in the context of digital financial services, consider a mobile network operator (MNO) that offers "air-time credit", or air-time loans, to its clients who normally pre-pay for services. Air-time credit means that when the subscriber runs out of money, the MNO immediately provides some additional air-time, and the client must pay for it within 10 days. If the air-time is repaid within 10 days, the MNO considers it a 'good' credit. Otherwise, it is a 'bad' or delinquent credit (i.e. repaid after 11 or more days).

Table 2a shows a simple way in which the differences in bad rates (column D) can be turned into scorecard points—by subtracting each bad rate from the highest bad rate (here 14.5% for men). The calculation and points are shown in columns E and F, respectively¹.

Table 2a: Scorecard Points Based on Bad Rates for Gender

	A	B	C	D (=B/C)	E	F
Gender	Goods	Bads	Total	Bad Rate	Calculation	Points
Male	3,877	657	4,534	14.5%	$14.5 - 14.5 =$	0
Female	5,123	343	5,466	4.1%	$14.5 - 4.1 =$	10.4
TOTAL	9,000	1,000	10,000	10.0%		

To build a scorecard, cross tabulations are studied for each potential risk factor. Those with bad-rate patterns that "make sense" are chosen for use in the model. Table 2b shows tabulations for Gender two more factors, Age and Days on Network².

Table 2b: Cross Tabulations for Gender, Age and Days on Network

	A	B	C	D (=B/C)	E	F	G
Gender	Goods	Bads	Total	Bad Rate	Calculation	Points	Max
Male	3,877	657	4,534	14.5%	$14.5 - 14.5 = 0$	0	
Female	5,123	343	5,466	4.1%	$14.5 - 4.1 = 10.4$	10.4	10.4
TOTAL	9,000	1,000	10,000	10.0%			

Age	Goods	Bads	Total	Bad Rate	Calculation	Points	Max
< 21	1,235	243	1,478	16.4%	$16.4 - 16.4 = 0$	0	
22 – 35	3,442	469	3,911	12%	$16.4 - 12 = 4.4$	4.4	13.5
36 – 50	2,447	231	2,678	8.6%	$16.4 - 8.6 = 7.8$	7.8	
>50	1,876	57	1,933	2.9%	$16.4 - 2.9 = 13.5$	13.5	
TOTAL	9,000	1,000	10,000	10%			

Days	Goods	Bads	Total	Bad Rate	Calculation	Points	Max
<90	1,433	406	1,839	22.1%	$22.1 - 22.1 = 0$	2	
91 to 180	2,140	295	2,435	12.1%	$22.1 - 12.1 = 10$	10.4	
181 to 270	2,354	203	2,557	7.9%	$22.1 - 7.9 = 14.2$	2	20.6
271 to 365	1,897	78	1,975	3.9%	$22.1 - 3.9 = 18.2$	10.4	
365	1,176	18	1,194	1.5%	$22.1 - 1.5 = 20.6$	20.6	
TOTAL	9,000	1,000	10,000	10%	$22.1 - 12.1 = 10$		

¹Many scorecard developers use more complex methods to transform bad rate data into scorecard points, but the results of the simplest method shown here and other methods will be similar.

²In the United States of America, indicators such as Gender and Age also are not used (along with race, religion) because of concerns about discrimination. When there is enough historic data on other relevant borrower characteristics and behaviors, it is indeed a good idea to leave out these factors—they are indicative on average, but wrong in many individual instances, which indeed has the effect of institutionalizing forms of discrimination.

The bad-rate patterns for these three indicators are:

- Risk is higher for men than women
- Risk decreases as age increases
- Risk decreases as time (in days) with the mobile network increases

Column G in Table 2b shows the maximum points for each scorecard indicator. The minimum points for each indicator is 0. This three-factor scorecard has scores ranging from 0 to 44.5 points. Three score-based risk groups are shown in Table 3.

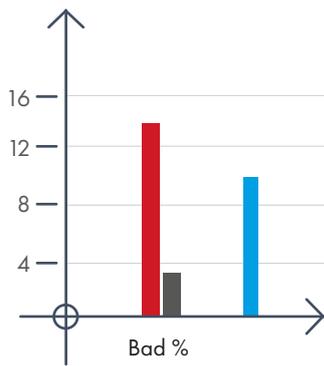
Table 3: Differentiate Borrowers by Risk Groups (or Total Scores)

Credit Score	>=	<=	Goods	Bads	Total	Points
Low risk	33	44.5	2,309	49	2,358	2.10%
Average risk	18	32	4,591	492	5,083	9.70%
High risk	0	17	2,100	459	2,559	17.90%
Total			9,000	1,000	10,000	10%

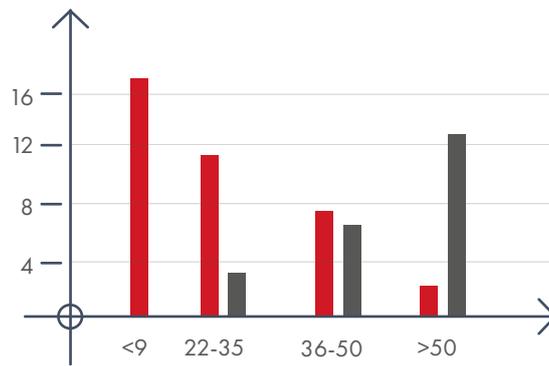
Summary: How Credit Scoring Works

Illustrations 1 and 2 on the next page present two hypothetical clients applying for an airtime credit. The model, with points derived from the past-bad rate relationships studied in the cross-tabulations, is used to 'score' each of the new applicants based on their gender, age and days on the network, and the average 'bad' rate for the group is the predicted bad rate for this new client.

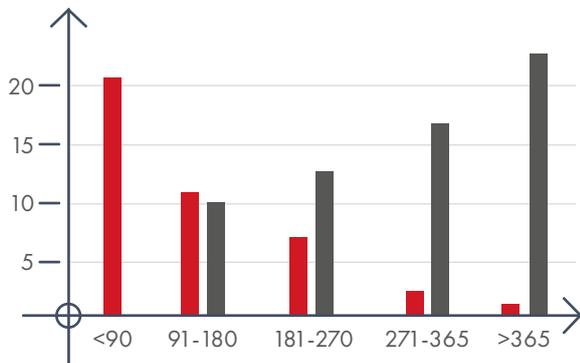
Illustration 1: Low Risk Client



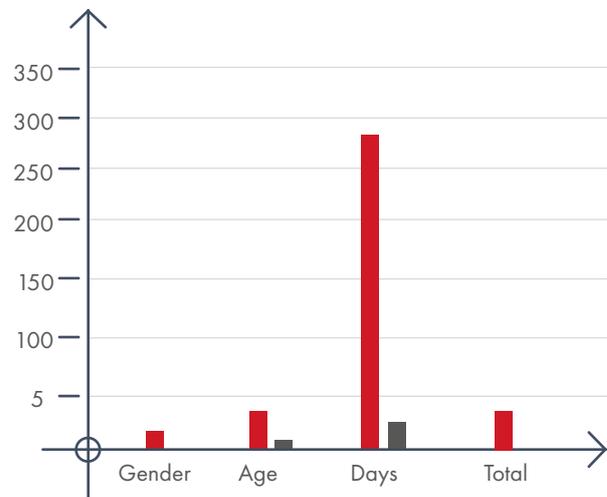
Female Male Points



Bad Points



Bad Points



Value Score



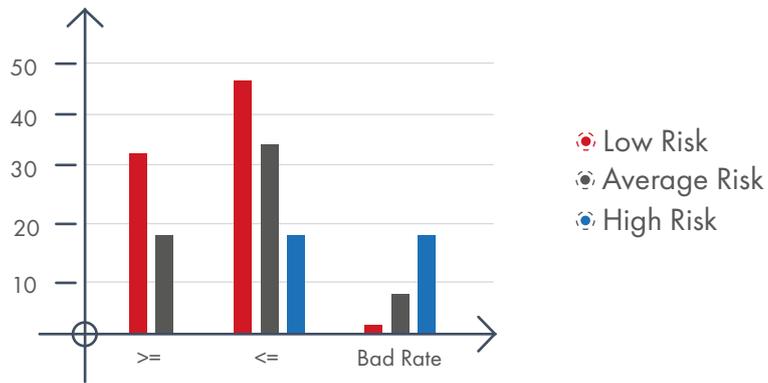
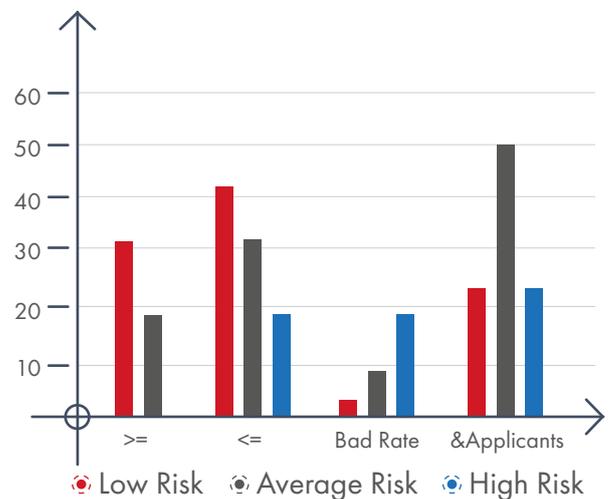
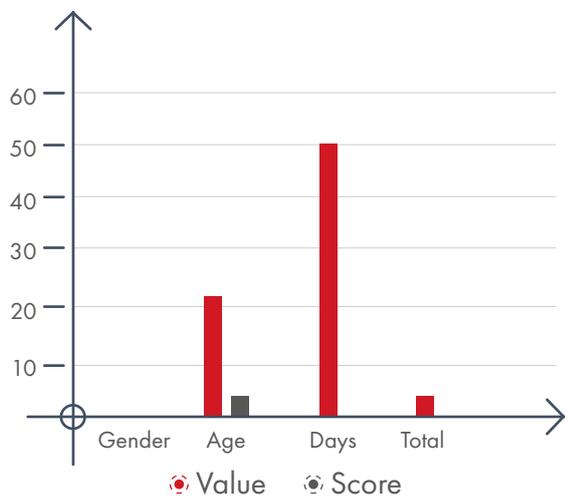
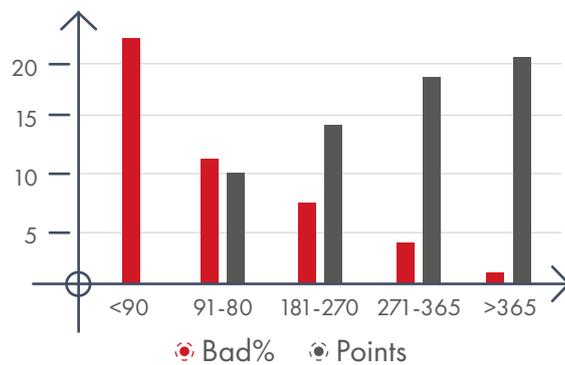
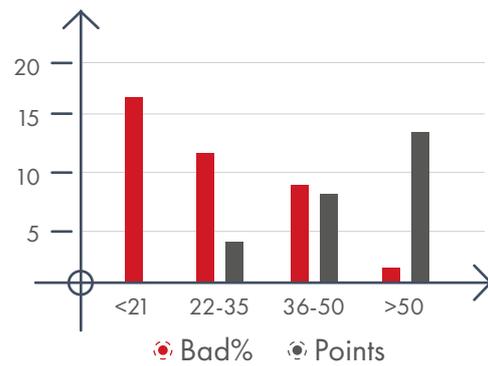
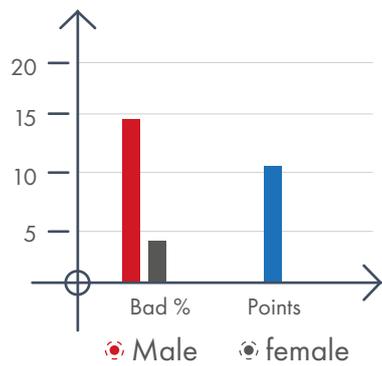


Illustration 2: High Risk Client



In the MNO's past experience, only 2 out of 100 (2%) customers like the lady in Illustration 1 (middle-age, some experience on the network) were delinquent on their air-time credits. By contrast, nearly 2 in 10 customers (17%) like the young man in illustration 2 (young and new to the network) were delinquent. This knowledge helps the MNO make appropriate decisions on loan sizes and interest rates offered to new applicants with similar risk profiles.

This example used only 3 risk indicators for simplicity. Actual scorecards use more indicators (i.e. 10 or more) to create a comprehensive risk profile of the borrower.

The next section looks at the types of digital data that are most likely to provide a comprehensive risk profile of small holder farmers.

What Kinds of Data Speak to the Risks of Small Holder Farmers?

When financial institutions offer digital saving and loans products to applicants they have never met, they rely on identity verification (so called "Know Your Customer", or KYC, data) and the digital data trail of spending on voice and mobile money services on a mobile network.

Identity verification and a digital services track record remain relevant for small holder farmers as individuals. In addition, a farmer's history of repaying personal loans (over digital or traditional channels) is strongly related to willingness to repay loans for agribusiness purposes. However, none of those data sources speak to a farmer's ability to plant, harvest and sell particular crops, or to breed, raise and generate income from animal. Other potential risks for farmers include timely input supply, soil and weather conditions, pests and diseases, reliability of off-takers, and potentially difficult-to-predict market prices.

This is where the data created by AFA program network service providers (and other multinational agribusiness suppliers and buyers) comes into play. A digital data trail of crop input purchases and/or sales objectively verifies a farmer's track record working with a given crop. A history of purchases and sales with other actors in a given value chain indicates the ability to honor agreements. Purchase and sales over digital platforms also document and verify cash flows that historically have been outside of the 'formal finance' system. Such digital transactional data, as well as other digital data that is directly relevant to farming (i.e. satellite and weather, e-learning) is explored in the next sections.

Purchase/sales records

Digital proof of purchases and sales verify at least a share of personal or business cash flow. Such data establishes a record of doing business over time and, also, sheds some light on affordability, or how much credit the business can afford given its cash flows. For example, AFA Network member iProcure captures dairy farmer input purchases on its digital sales platform. This gives an idea of each of its clients cash flows and track record (based on the length of time the farmer has been active on the iProcure platform) in dairy production. The potential uses of data collected by Twiga Foods, another network member, are presented in the call-out box.

³Concept borrowed from Naeem Siddiqi (2005) Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. Wiley

TWIGA FOODS

Twiga is a mobile-based, cashless business-to-business supply platform for Africa’s retail outlets, kiosks, and market stalls. Vendors order stock from Twiga and receive it at their shops on the next day at a price and quality that usually compares favorably with informal markets.

Twiga was recently collecting 12 types of transactional data—4 related to deliveries and 8 related to collections (purchases). Each data field's potential relevance to credit risk is presented in Table 4.

Table 4: Twiga Data Fields and Potential Relation to Credit Risk

Ref	Description	Potential Relevance to Credit
RETAILER PURCHASES		
1	Amount paid for delivery (purchase)	Evidence of cash flow (affordability)
2	Location of delivery	Different purchasing power in different regions
3	The route the delivery was made on	
4	The team member making the delivery	No likely relevance to credit risk
FARMER SALES		
1	Amount paid to farmer	Evidence of cash flow (affordability) and sales track record
2	The area the harvest was done in	Geographic risk
3	Location of collection	
4	Farm where the collection was harvested *	Identification of the farmer
5	The product that was harvested	Crop risk, which can include risks arising from price volatility, weather risks, etc.
6	Weight (in kg) of the harvest	
7	The unit price we purchased the product at	
8	The individual weights of produce that sum to "quantity"	

⁴For more information see <http://twigafoods.com/>

Cash flows documented on Twiga's platform could definitely help financial institutions determine how much a farmer or stall owner can afford to borrow. It would need to study the other indicators in relation to repayment of past obligations to understand how and if they are indeed related to credit risk.

E-learning on relevant topics

Digital e-learning on agriculture and finance-related topics could potentially contribute to positive outcomes for farmers borrowing in the formal financial sector for the first time. Relevant e-learning content could also include timely updates about weather, market prices, and networking or educational events/opportunities happening in the community.

Data collected on e-learning platforms is potentially useful in predicting credit risk. While there is still much practical and research to be done in this area, one promising service e-learning service provider is Arifu, discussed in the box below.

ARIFU

Digital service innovator and AFA Network member Arifu provides free and interactive distance learning over mobile phones. It has offered digital learning content on directly relevant topics including agronomy, entrepreneurship, financial literacy. Its software platform captures a wealth of behavioral data on farmer's engagement with the service, including their e-learning's paths and outcomes.

Arifu currently collects data on:

- **Courses taken/depth of engagement:** the quantification of learner engagement on topics like agronomy, inputs, and financial products may confirm expectations that better knowledge of how to use loans and ag inputs reduces credit risk.
- **Demographics:** self-reported and inferential logging of gender, age, income level, and education level can corroborate KYC information to verify identity. Such demographic factors usually demonstrate somewhat 'universal' relationships to credit risk (ie credit risk decreases as age increases, credit risk for small loans decreases as income increases, women are lower credit risks than men, etc.)
- **Psychometrics:** Arifu is working with psychometricians to identify users' e-learning behavior patterns. They hope to map usage patterns that are conceptually similar to studied personality qualities or 'traits' that may be related to willingness to repay a loan—for example, 'grit/perseverance', 'conscientiousness', and 'locus of control. Inferring personal qualities from behavioral data, rather than directly asking personality test questions, has the potential to reduce the 'noise' in such data that results from language, answering styles, and situational behavior (but its efficacy has yet to be tested).
- **Other:** through some sub-applications, Arifu can capture info on a subset of learners such as their location (sub-county level), self-reported goals (life, business, savings goals, etc.), products of interest, farm size, level of engagement in content/application, etc. Data points such as these may also be useful in ranking borrowers by credit risk.

These new types of data need to be studied together with loan repayment to find which types work best and are complimentary (or redundant) to data traditionally used in credit scoring models.

Farm-Related Locational data (GPS, satellite)

Knowing the precise location of a farm without visiting it in person makes it possible to use other ag and weather data sources to understand things such as:

- Soil type and quality
- Expected rainfall
- Water and pest risks in the area

To use ag-specific data for credit risk assessment purposes, it is necessary to know a very accurate location of a farm and pair that with the weather and soil data for that same location. It also necessary to have expertise on the soil and weather conditions conducive to the types of crops under production in order to understand potential risks.

Table 5 is an example of a look-up table for required temperature ranges, water needs, and mean farmer income per acre for 5 crops. With such tables, it is possible to systematically check that the crops a farmer produces or claims to produce is feasible for the farmer's verified location.

TABLE 5, TEMPERATURE AND WATER NEEDS PER CROP

	TEMPERATURE RANGES FOR CROPS (in C)			WATER NEEDED (MM Growing Period)	MEAN INCOME (p/Acre)
	Optimal	Max	Min		
CORN	22-25	32	20	-	
WHEAT	20-25	38	5	450-650	
RICE	30-33	37	22	450-700	
POTATOE	15-20	28	12	500-700	
SOYBEAN	25-28	37	14	450-700	

Checks of farmer data against benchmarks and norms of the type in Table 5 do not need to be part of a risk-ranking model. Instead, they can be encoded into checklists or 'business rules'. Some example business rules could be to reject (or flag for some additional analysis) cases where:

- Temperature in next 10 days outside norms for crop
- Projected rainfall outside of norms for crop
- Farmer Income less than 70% of median farmer income for crop
- One or more crops under production does not fit within soil condition norms in farm location

The Potential Value of Self-Report Ag-Related Data

Farmers also can directly provide information about their activities, such as:

- size of the land under cultivation
- type of crops planted
- how often the crop is planted
- years experience planting a given crop
- expected (or past average) yield
- expected sale prices

This is the type of information micro-finance loan officers ask for when visiting farms seeking finance. Even without specialized knowledge about the cultivation of a given crop, the loan officer can visually verify that the farm and appropriate farming equipment exist. Perhaps most importantly, the loan officer can check with the local community about the farmer's reputation—how long has the farm been in the community, has it been successfully growing and selling its produce, etc.. This is standard community, character-based lending, and many lending decisions for first-time borrowers end up being based on trust engendered by personal contact.

Digital channels are potentially much less costly and time consuming than personal visits by loan officers, but they offer less scope to verifying self-reported information from the farmer. This is why credit scoring-based solutions for digital financial services would work best with verifiable digital data from known digital service providers. In such cases, there is no need to ask a farmer how much was purchased or sold – actual purchases and sales behavior is evidenced by data collected on the digital purchases or sales platform.

Potential Use of Other Non-Ag Related "Alternative" Data

Some vendors have developed proprietary models based on other types of data that can be collected digitally. Two types that have received media attention are psychometric testing and use of social media information.

Psychometric Testing

Psychometric testing has the potential to provide a check on a loan applicant's character, much as a loan officer would attempt to do through personal contact and 'reference checks' with business

partners and neighbors. However, unlike traditional lending character checks, which are based on personal observation and independent references, psychometric testing relies on answers to questions that have been specially designed to measure particular personality traits. While psychometric testing has a long and successful track record of usage in human resources and clinical psychology, the accuracy of its personality measurement in lending situations is largely unstudied and unknown.

It is possible to rank borrowers by risk based on answers to personality test questions, but the method is subject to several challenges to data consistency and 'traditional' lending processes, including:

- People understand questions differently based on their backgrounds
- People answer questions differently in different situations
- Personality tests have not traditionally been a part of the loan application process

Social Media Data

Social media data and mobile-device data (such as number of contacts, and content of text message and emails) has reportedly been used successfully by some fintech start-ups to verify identity and, perhaps to a lesser extent, rank borrowers by risk of repayment for the purpose of consumer lending.

The number and nature of a person's social media contacts likely says something about that person's 'stability', as an individual or in a community. However, people use social media for various reasons and to various degrees, and again these differences and variations make such data less likely to provide a consistent indication of likelihood of repayment in comparison to more directly relevant things, such as business track record, payment of other obligations, membership in a cooperative, etc.

As explained in the next section, the more available data that is directly relevant to agribusiness, the less likely other alternative data sources will make material improvements to prediction. This may focus the efforts of technical assistance programs and fintech start-ups targeting small holder farmers on collecting more and better directly relevant data over digital platforms in order to unlock credit in cooperation with existing financial institutions.

Putting it All Together

To summarize, there are various types of data that can be collected over digital channels. Their value to prediction in credit scoring models is likely a function of:

- The cost of collecting the data
- The extent to which the data is available from all applicants
- Relevance of the data to a farmer's ability or willingness to repay a loan

Given a set of data, the risk ranking power of a multi-factor credit scoring model increases as each individual indicator that ranks risk well is added to it. At a certain point, however, (which can be as few as 10 indicators for some types of borrower segments) additional factors will add little to the model's risk ranking. This means thoughtful consideration of what data will be collected is likely to lead to better results for credit scoring models. Digital data about purchase/sales behavior, usage of credit products, and usage of other digital services (voice and mobile money) have the benefit of being verifiable and directly relevant. As digital service platforms gain popularity with small holder farmers, the opportunities to measure their risks and offer appropriate financing will increase.

Data Collection Strategies to Facilitate Scorecard Development

No matter what type of data a service provider has collected, it will also need data about repayment of past obligations to develop a credit scoring model. As mentioned in the background on credit scoring models, each piece of data is analyzed to look for characteristics that better differentiate between good and bad loan repayment.

How can a digital service provider that does not offer credit match its own data to loan repayment data in order to develop a model? Some possibilities include:

1. **Consent to Access Credit Bureau Reports:** In much of Africa, the quality of credit bureau data has significantly improved over the past few years. Most notably, the inclusion of data on digital savings and loan products in Nairobi's three commercial credit bureaus has greatly expanded the number of people with a credit track record from formal financial institutions. Asking digital service users for consent to check their credit report (for example, in order to receive future offers for credit products) allows credit bureaus to calculate and sell retroactive credit scores (or scores as of a past date) for the digital service provider's clients. The provider could then analyze which types.
2. **Partner with a Creditor:** If there are large creditors in the market who work with the same client base as a digital service provider, it can explore the possibility of working together to develop a scoring model. The partner could be a bank or MFI that has provided formal bank credit, or, as in the case of M-Shwari, the consumer loan product, could initially be an MNO that offers air-time credit. The partner could be also a digital service provider that offers invoice discounting or trade credit—the key is that the partner would have some data to segregate clients into 'good' and 'bad' payers to study risk relationships for a credit scoring model.
3. **Pilot Products with Risk-Limiting Conditions:** another strategy is to offer assume that certain qualities reduce initial repayment risk and use a conservative strategy to offer first loans to clients that are have these qualities. For example, small first loans can be offered to:
 - buyers/sellers meeting some threshold-level of activity on the platform
 - clients with past loans and no serious delinquencies
 - cooperative membership for over 5 years with no negative history

Such strategies have long been used by international credit card companies to build credit history for students and employees. Those studying or with steady employment are given a card with a small credit limit, and limit increases are offered after a sufficient period of successful use and repayment of the credit card. As digital service platforms grow in prevalence, financial institutions will have new channels through which to help farmers build credit history and, potentially, graduate from simple starter loans to access longer-term and/or more complex financial products and services.

Working with Commercial Partners

This paper has emphasized the importance of digital data itself and how its consistent creation, collection and analysis can help financial institution better assess the credit risks of farmers and help them to build credit histories that will gradually unlock the financing they need to grow and prosper to their full potential.

There are also several areas of sensitivity and concern that digital service providers and their commercial partners need to take into consideration:

Consent, Data Privacy and Fraud Protection

Farmers, like all of us, are concerned about how and by whom their data will be used. For this reason, it is important to ask for and receive the necessary informed user consent at the time of client onboarding and expansion of services. Particularly important for scorecard development is asking consent to obtain a client's credit bureau report, but more generally, clients must be informed and agree to how data they provide may be used in the future.

Business Requirements/Service Level Agreement Terms

When working with third-party suppliers of data or data analysis services, it is important to define clear and actionable business requirements and service response levels. If a scoring model depends on timely data from a weather service or MNO, the partners must agree in advance what will happen in event of any technical disruption of service—particularly how financial losses are shared.

Conclusion

Many types of data now being collected from smallholder farmers have the potential to inform the assessment of their creditworthiness, including, but not limited to:

- › Digital data track records of time in business and past purchases and sales
- › Successful completion of e-learning/training content related to crops under production
- › Any information on repayment of past obligations (digital loans, loans, utilities, supplier credit, etc.)
- › Other spending documented on mobile networks (air time and money)

It turns out that the data sets being built for entirely different reasons—to trace the path of a crop to market, to teach a farmer how to improve his yields, to provide farmers with timely, price-competitive inputs—contain the same information banks want to see when deciding if a farmer will be able to repay a loan.

The challenge is to pair the agri-data with financial partners and other stakeholders willing to take some initial risk to unlock future rewards—namely, opening a large, underserved market for financial institutions while potentially improving the productivity and livelihoods of millions of small-holder farmers.



DIGITAL FINANCIAL SERVICE FOR SMALL HOLDER FARMERS

What data can financial institutions bank on?

November 2017



**MERCY
CORPS**

**AGRIFIN
ACCELERATE**

AgriFin Accelerate is a six-year initiative supported by the **Mastercard Foundation**, working in Kenya, Tanzania and Zambia.