MAKERERE UNIVERSITY
COLLEGE OF COMPUTING AND INFORMATICS TECHNOLOGY

DATA KIT SYSTEM TO INCREASE ON THE EFFICIENCY OF DATA COLLECTION AND ANALYSIS IN UGANDA

(Case Study of Micro-Finance Institutions)

By
CS19-13
DEPARTMENT OF COMPUTER SCIENCE
SCHOOL OF COMPUTING AND INFORMATICS TECHNOLOGY

A Project Proposal Submitted to the School of Computing and Informatics Technology
for the Study Leading to a Project in Partial Fulfillment of the
Requirements for the Award of the Degree of Bachelor of
Science in Computer Science of Makerere University.

Supervisor

MS. BARBARA NANSAMBA
Department of Computer Science
School of Computing and Informatics Technology, Makerere University
bnansamba@cis.mak.ac.ug, +256-77-494-5928 ,

#	Names	Registration Number	Signature
1	ABILA RAPHAEL	16/U/2673/PS	
2	MWAITA JOSHUA	16/U/7890/PS	
3	OJIAMBO ABEX	16/U/10829/PS	
4	NAKAFEERO PENINAH	16/U/8463/PS	

APPROVAL

This Project Proposal has been submitted for examination with the approval of the following supervisor.

Signed: ..... Date: .....

Ms Barbara Nansamba
BSc. CS, MS.DSE School of Computing and Informatics Technology.
College of Computing and Information Sciences Makerere University.
bnansamba@cis.mak.ac.ug.

Contents

List of Figures

List of Acronyms

#	Acronym	Meaning	
1	ODK	Open Data Kit	
2	GPS	Global Positioning System	
3	UML	Unified Modelling language	
4	ASD	Agile Software Development	
5	OXD	Open X Data	
6	HTML	Hypertext Markup Language	
7	CSS	Cascading Styling Sheets	

1 Introduction

1.1 Background

Over the recent years, Research has been used as a basis to make important decisions in most institutions of Uganda. In order to plan and manage resources, public institutions need accurate and timely data(Deo, 2012). Many institutions have experienced a high cost in terms of time and money when it comes to this area simply because of the poor techniques of data collection that are being employed.

Given the fact that a successful research must result into accurate information that must be reliable when making important decisions; some research has ended up being unreliable and this is because most of the institutions take it for granted and have resorted to estimating the results of their research. All secondary data is derived from the primary sources and it is often difficult to track the origin from source and whether the secondary data has been replicated, aggregated or modified from the original. Only 29 percent of datasets published by international agencies are derived from primary sources.(Wilbrod Ntawiha, 2016)

There has always been a high risk of errors in the output of the results, majorly attributed to the current unreliable systems that are being used to gather and analyze data as researchers in most institutions try to collectively gather their results and input them manually to their website servers from tools such as mobile phones, flash disks, tablets and papers. These approaches also deny the administrators the chance to monitor the work being done by their researchers since they have no guarantee that the research was carried out in the right place by the people deployed to carry out these research activities.

With the advancement of mobile technology, the demand of advanced and mobile phone based methods of collecting data have increased and therefore researchers have been devoted to study how this technology can improve the data collection process. (Deo, 2012)

1.2 Problem Statement

The continuous delay in the research carried out in most institutions in Uganda is attributed to the poor techniques of data collections and analysis. Currently the data is recorded by using hand-held computers or specialized data recorders which is then transferred to more secure devices or servers. This does not only involve a lot of time and labor but also contributes to the loss and inaccuracy of data during the data transfer and analysis since there is a possibility of people manipulating this data in favor of their desired persons that they want to benefit from the outcome of the research. Some researchers end up with either incomplete results or no results at all due to unreliability of the existing system and poor monitoring strategies by the administrators. This calls for the development of a data kit android application which will be used by researchers in data collection, it will synchronize the data to the administrative website with the location of the research automatically where analysis of the data can be performed and through the analysis, it will be able to recommend areas that need assistance or more research. This will improve efficiency in data collection, employee management, decision making by the board as well as reduced costs involved in research.

1.3 Objectives

1.3.1 Main Objective

To develop an automated online research tool system that will increase the efficiency in the process of collection and analysis of data by micro-finance institutions and to recommend on the areas in which more research is required.

1.3.2 Specific Objectives

The specific objectives of this project are to:

- To establish requirements to the real time research tool system.
- To design the android and web application sections for the Data kit system.
- To implement the designed android and web application section for the data kit system.
- To test and validate the Data kit system.

1.4 Scope

The study is focused on producing a real time research tool to help micro-finance institutions in gathering and analyzing information about the financial status of peasant farmers in Uganda so as to offer them loans to improve on their agricultural practices thus boosting the Agricultural sector financially.

It will also focus on helping the micro-finance institutions to monitor the research carried out by their employees.

The study will target at covering the farmers in the whole of Uganda however the number of farmers in Budaka district will be used as the sample size for the study.

1.5 Significance

The new system will provide a synchronization between the android application and the website server hence providing timely updates of the data collected from the fields. This will also help to reduce chances of errors in the data collected as well as manipulation of this data.

It will also automatically analyse data uploaded to the server and will save a lot of time wasted in the data analysis by the researchers themselves.

The new system will help micro-finance institutions in recommending areas that need more research basing on the outcome of the data analysis. This will increase on the efficiency of research since researchers will only be left with the task of gathering data.

Given the fact that this new system will collect the GPS(Global Positioning System) longitude and latitude coordinates of the geographical area where research is being carried out, the administrators will be able to monitor their researchers so as to get rid of data estimations which are currently being practiced by target workers in most institutions.

The new system will increase on data integrity. This is because the data collected will be stored on a robust data system that cannot be accessed by intruders. This will reduce on the unwanted manipulation of data.

2 Literature Review

2.1 The Current Research system

The following are the current systems that are in use for the data collection and analysis: -

2.1.1 Open Data Kit tool application

Open Data Kit (ODK) is a suite of tools that allows data collection using mobile devices and data submission to an online server, even without an Internet connection or mobile carrier service at the time of data collection. You can collect data remotely without an Internet connection or cell carrier access. Gather text, numeric data, media and more with a mobile device. Then, host your data online using Google's powerful hosting platform, Google App Engine, and visualize your data as a map using Google Fusion Tables and Google Earth. (Earth, 2015)

The challenge with this tool includes the following: -

Errors in the time and date of research. These errors are due to differences in date settings on mobile devices as well as an absence in start time thus the need of using medium-to-high grade Android phones so as to eliminate errors such as those experienced with date and time on some of the phones used.

Poor usability. Open Data kit is difficult to set up . This is because it involves a lot of stages to create the data collection platform and this may affect researchers who may not have the technical skills to operate the software required to setup the over all system. Not only that but also ODK has poor visualization interface that may be difficult to work with.(Earth, 2015)

2.1.2 Manual data collection and entry system.

This is the most widely used system of data collection. it involves entry of data collected manually from data collection tools such as phones, paper, and so on. This system comes with a lot of challenges such as: -

Too Much Money. Manual data entry takes a lot of money to be conducted efficiently. Employees need to be properly trained for any kind of data entry and training is a huge amount of time and money for the company.

Human Error. The next problem in relation to data entry is human error. Everyone knows that humans are not perfect and there are going to be times when mistakes are made, whether it is spelling, grammar, or punctuation. Using automation will reduce the margin of error.

Time Consuming. When entering information manually, it takes up a lot of time, no matter how fast one can type or process information, it will never be as fast as it needs to be. Typing in a lot of information can be tricky and cause individuals to lose focus and this will also add to the time it takes for data entry. Not being fast enough leads to a big delay in data availability.

Misinterpretation. The last issue with data entry is misinterpretation; this means that because the entries are being conducted by humans there is always room for misunderstanding. Everyones mind thinks differently and understands things in different ways and the same goes for data entry.(Biel, 2015)

There are many problems associated with manual data entry but there are also many ways to combat these problems through application of advanced and upcoming technology.

2.1.3 Open X Data

Open X Data(OXD) is an open source data collection framework, which has been championed by the community of academic researchers and developers from various universities across the world. Open X Data support low-end java enabled phones to collect data remotely in low budget settings. It provides a way to visualize and manage the gathered data though a web interface. (Deo, 2012)

2.2 The Intended Data Kit Application.

2.2.1 Comparison with Open Data Kit (ODK)

The newly intended data kit application system will analyse the data collected by the researchers in the field and yet this functionality is missing the Open Data Kit.
With The help of machine learning technology, the newly intended data kit application will provide a timely recommendation of areas that are pending in terms of research which is still missing in the Open Data kit.
The Open Data Kit is poor in terms of usability because it involves a lot of technicalities in terms of setting up the platforms for data collection and this makes it hard for people with little Technical skills to work with it. Therefore, the new system will over come this challenge by only allowing the users to use a simple mobile application to download forms from the server without going through a lot of stage of setup.

2.2.2 Comparison with the manual system.

The newly intended data kit application will overcome the challenges encountered in the manual system in the following ways: -

1 The newly intended system will provide automation in data analysis that is absent in the manual system.

2 The newly intended system will have system security that is absent in the manual system hence unwanted data manipulation will be avoided.

3 The Data kit application system wil alsol reduce the costs involved in data entry because data collected from the field is automatically uploaded to the server website. This will overcome the error challenges faced in manual data entry.

A Table showing the comparison between the New system and the existing systems

#	SYSTEM	Recommendation	Data analysis	errors	GPS location
1	Manual System	No	No	Yes	No
2	Open Data Kit	No	No	No	Yes
3	Open X Data	No	No	No	No
4	The New Intended system	Yes	Yes	No	yes

3 Methodology

3.1 Introduction

This chapter comprises of a detailed description of selected methodology of how we intend to achieve the objectives of the study/research through step-by-step methods and techniques of data collection to get system requirements, system design, implementation plus testing and validating the implemented data collection kit.

3.2 Research design

A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure.(kothani, 2004) Research design that will be used is the sampling design through the selection of the different individuals from a multitude to represent other people and the result of research will be generalized to the population from which those individuals will be selected.
The research will be based on system development life cycle stages like analysis, design, development, test and implementation will be followed throughout the development of the Proposed data collection kit.

3.3 Sampling

The random sampling method will be used and this involves selecting a sample from a population under study.This will reduce bias in selecting sample size.
The population size under consideration is the number of farmers in Budaka district.

3.4 Development approach

The agile development methodology will be used in the development process. ASD will help us in keeping simple, testing often, and delivering functional bits of the application as soon as they are ready. This will help us to minimize risks such as bugs, cost overruns and changing requirements. The Github platform will be used for collaboration among the different members of the group. ie each member will upload their code on the Github platform for others to view.

3.5 Methods of Data collection

In order to identify the requirements of the Real time Research tool system, data shall be collected and analyzed. Data collection and analysis techniques that will be used include the following: -

3.5.1 Interviews

An interview is a data collection method where a researcher asks a respondent a set of questions and records his/her answers. we will be in position to interact with different people face to face using an organized interview guide that will be in the line with the rest of other tools of data collection to be used elsewhere during the study. This will also be a building block of the system requirement collection.

3.5.2 Questionnaires and Surveys

A questionnaire is a research instrument consisting of a series of questions (or other types of prompts) for the purpose of gathering information from respondents. This technique will help us to collect data from a large group of people that are scattered. some of the data collect will include opinions of individuals about the current systems used to collect data from the fields. With this in place, we shall have known how to develop a system with a good user experience.

3.5.3 Observation.

Observational research is defined as the method of viewing and recording the actions and behaviors of participants. This will include going to fields of study.
Observation will help us to confirm certain aspects so as to ensure evidence of claims made by interviewee.

3.5.4 In depth literature review

This is the analysis of the existing documentation on a given object. We shall use this approach to find out how the related applications operate by looking at their strength and their weaknesses to help us find ways of making our project better in terms of project implementation.
This will contribute to the knowledge that we currently have with the existing tools and systems used for data collection and analysis. Several document previews are to be continuously conducted throughout the period of the study about the existing data collection systems of different organizations and this will consist of visiting books, several web portals of different organizations, journals, different reports directed in the same line of study, internet among others in search for vital information that is to be used throughout the course of the project.

3.6 Data Analysis

The data collected will be analyzed using both qualitative and quantitative methods in order to generate information from the meaningless data for decision making which will help in planning, design and implementation of the proposed system. Quantitative methods will help to realize the number of formally used data collection and analysis systems. Qualitative methods are to help us analyze different already existing data collection systems. We shall use the stastic package SPSS to analyse the quantitative data collected.

3.7 System analysis and design.

This will be achieved using Unified Modelling Language (UML) to come with the different designs of the system like use case, activity diagrams and so on. This new system will be designed and described using the following UML objects: -

3.7.1 Class Diagram

We shall use this diagram to describe the attributes and the constraints imposed on the system.This model will be used because our system is object oriented and Class diagrams are the only UML diagrams, which can be mapped directly with object-oriented languages.

3.7.2 State Diagrams

We shall use state diagrams to describe the behavior of the system considering all the possible states of an object when an event occurs. This behavior will be represented and analyzed in a series of events that occur in one or more possible states.

3.7.3 Use case Diagram

This is used to represent the different instances of system use by the different users of the system.

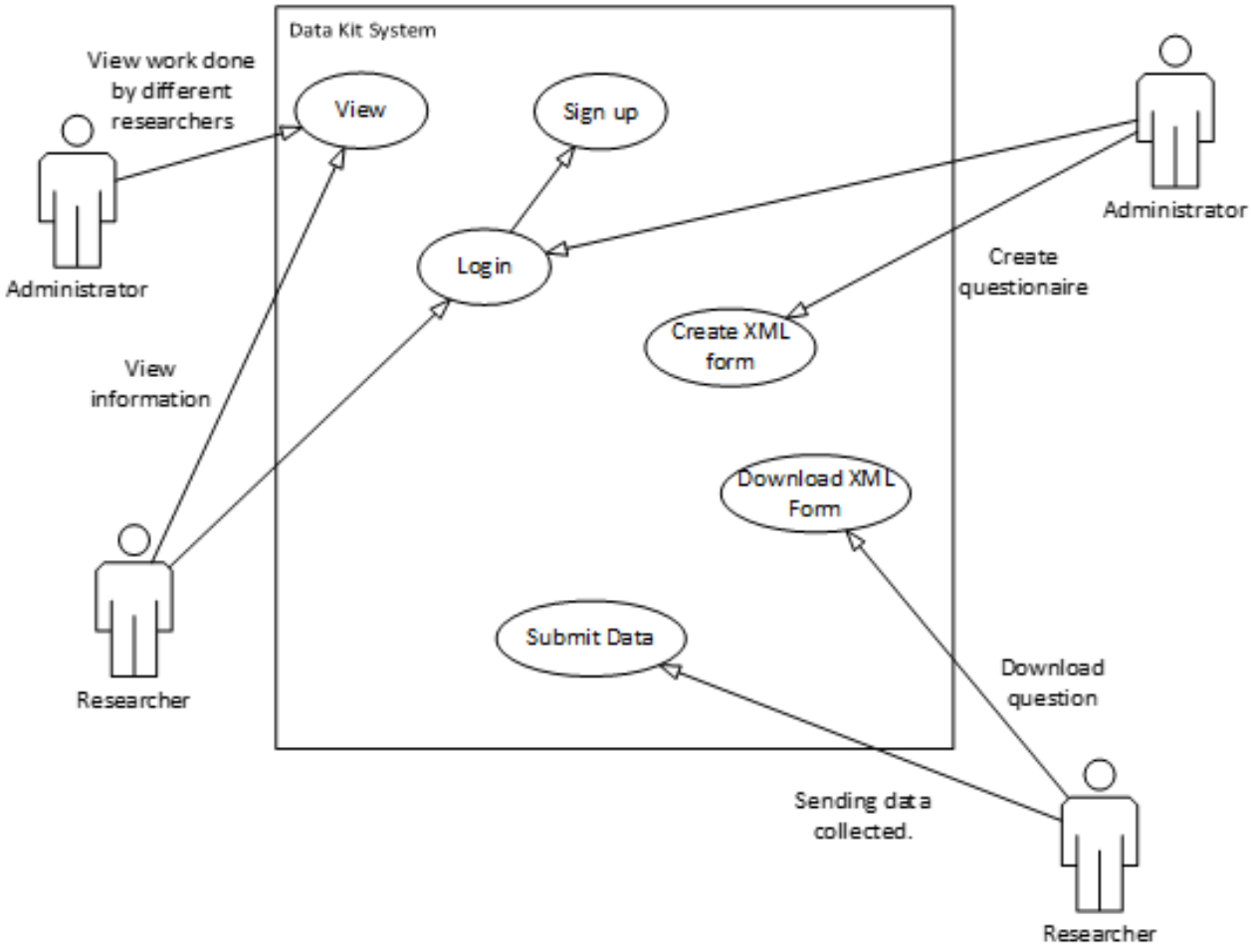


Figure 1: Use case diagram showing how each user interacts with the system

3.8 System Implementation.

This is a stage where visions, plans become a reality. It's a logical conclusion, after evaluating, deciding, visioning, planning and applying for the proposed project. This process will involve having the actual plan in place, achieving tangible change and improvements from one stage to another, ensuring that any unforeseen conflicts that may arise are neutralized before continuing to the next phase.

3.8.1 System Architecture

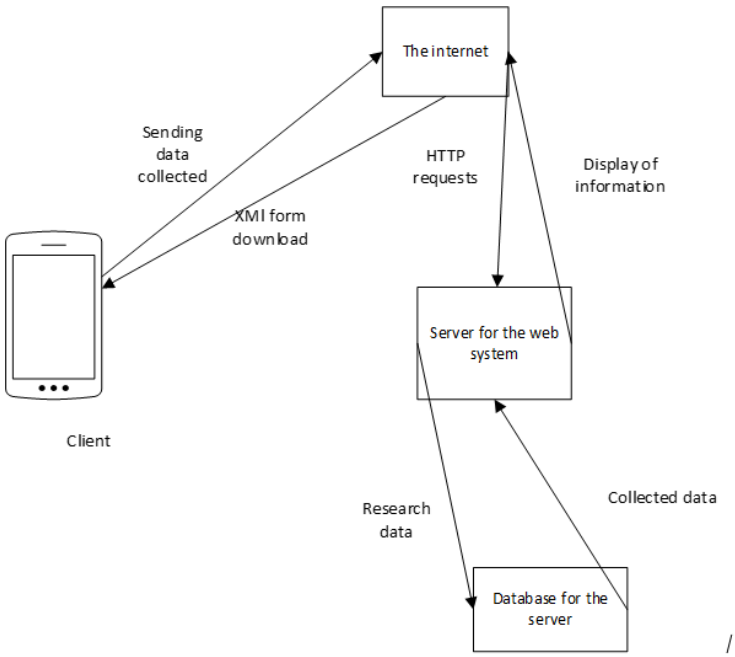
The newly intended system will run on a client-server architecture.

Client-side

This is the android application that will be connected to the web server. This is application will be used to retrieve XML forms from the server using a mobile phone connected to the internet. The researchers are to use this application to send the data collected to the web server.

Server-side

The server will receive data sent from the android application and carry out analysis .
Below is a diagram showing the Architecture of the intended Data kit tool: -



3.9 Tools to be used

3.9.1 Android studio

This is a tool that will be used to produce the android application and it uses both Java and XML. XML is a markup language just like HTML and with the help of this Language, we shall develop

the interfaces of the android application. Forms of the data kit system will also be created using XML.

3.9.2 Hyper Text Markup Language(HTML)

This is the standard markup language that is used to create the structure of the different components of the websites such as tables, forms, color of the different elements etc.

3.9.3 PHP language.

The application database will be linked using PHP scripting language. PHP will also be used to develop scripts that connect and pull data from the database and display it in the system.

3.9.4 Javascript

JavaScript is a client- side scripting language that will be used because of its ability to handle form processing burden. JavaScript also promotes dynamism in the system

3.9.5 Bootstrap

Bootstrap is the most popular CSS Framework for developing responsive and mobile-first websites. This comes with already built CSS for improving the styling the elements of the website.

3.9.6 MySql

This is a relational database management system that is open source and free, runs on all platforms available and its able to handle large data manipulations. This will be used in creating relational data bases of the system.

3.10 System Evaluation and Testing.

This is the process of executing the application programs with the intent of finding errors and observing if the program is behaving as expected. Testing of the program will be first done by the developers and then taken to the users for further testing. The following testing Techniques will be used: -

Unit Testing. This involves the testing of the individual unit of the code to check if they are fit for use. A unit is the smallest testable part of an application.

System testing. This is a complete replication of the running system for the purpose of testing out the adequacy of the system.

Compatibility Testing. This will be done to ensure that the system is compatible with the existing phones and the protocols under which data is transmitted.

References

Biel. (2015). *Automation and efficiency.* Retrieved from http://biels.com/the-problems-of-manual-data-entry/

Deo, S. (2012). A proposal of a mobile health data collection and reporting system for the developing world. *Master Thesis Project*.

Earth, G. (2015). *Mobile data collection using the open data kit.* Retrieved from https://www.google.com/earth/outreach/learn/mobile-data-collection-using-odk-collect/

kothani, C. (2004). *Research methodology methods and technology.* New age International Publishers.

Wilbrod Ntawiha, B. A. (2016). Ugandas data ecosystem. *A quantitative review of data production in Uganda*.

4 Appendices

NUMBER	ACTIVITY	COST(Shs)
1	Gathering information	150,000
2	Design a framework	100,000
3	Implementation	200,000
4	Miscellaneous	200,000
	TOTAL	650,000

Figure 2: project budget

	Task Mode	Task Name	Duration	Start	Finish	Predecessors	% C
1		group formation and problem identification	1 day	Tue 8/28/18	Tue 8/28/18		100
2		submission of the concept paper	9 days	Wed 8/29/18	Mon 9/10/18	1	100
3		data collection and analysis	14 days	Tue 9/11/18	Fri 9/28/18	1,2	50%
4		system design and modelling	14 days	Mon 10/1/18	Thu 10/18/18	1,2,3	90%
5		Proposal writing and submission	20 days	Fri 10/19/18	Thu 11/15/18	1,2,3,4	95%
6		Implementation and prototyping	21 days	Fri 11/16/18	Fri 12/14/18	1,2,3,4	30%

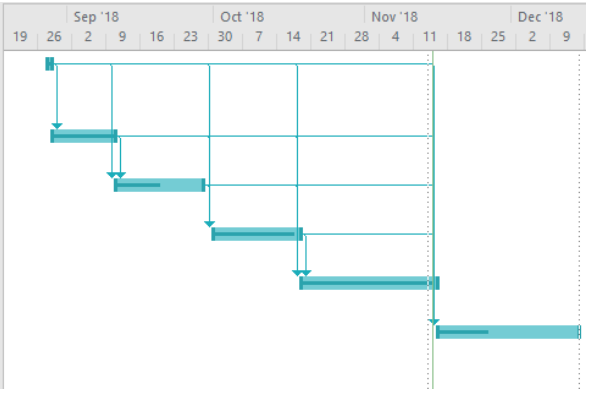


Figure 3: gannt chart showing project schedules

MAKERERE UNIVERSITY

P. O. Box, 7062,

15th Nov, 2018.

UGANDA MICROFINANCE LIMITED

KAMPALA

Plot 49/51 Bukoto Street

P.O.BOX 10184

041 4259176

Dear sir/madam

RE: PERMISSION TO USE YOUR INSTITUTION AS A CENTER FOR RESEARCH

I am a Makerere University year three student who is pursuing a bachelor of science in computer science

On behalf of my colleagues would request you to offer us permission to use your institution as a basis for our research.

We are carrying a research to collect data that will guide us to develop a data collection kit that will aid data collection and analysis for loan offers by microfinance institutions (Credit evaluation).

We therefore seek your permission to use your institution as a center of our research.

It will be of great offer if our request is put into consideration in your office.

All information provided will be treated strictly as confidential and purely for academic purpose. Looking forward to your favorable response.

Yours Sincerely,

Mwaita Joshua

Figure 4: letter of approval

The following are the places to be visited

Place	Address	Contact
Pride microfinance	Nakulabye (Namirembe Road)	041 4345709
UGAFODE Microfinance Limited (MDI)	Rubaga road	041 4257181
UGAFODE Microfinance	Bombo road	0414257181

Map showing location of some microfinance

Figure 5: list of places to be visited

CREDIT EVALUATION FORM FOR MICRO FINANCE INSTITUTE

Bio-Data CLIENT BIO-DATA

1. Capture the interview date. Client name Client Age (in years)

2. Client Marital Status Married:

a. Widowed b. Separated c. Divorced d Single

3. Client GPS locations coordinates can only be collected when outside.

4.Loan requested for in shillings Client Telephone (+256)

5. Client District Client Sub-county Client Village

STATRT OF CAPITAL ENDOWMENT

1. Do you know your neighbors Yes 18 No 0

2. Do you employ people? Yes No

3. List the two people who can guarantee your loan?

START SELF WORTH FORM

1. Do you have assets like land, house or animals? Yes or No
2. Do you have land.? Yes or No
3. Take a Screen Shot of the Land Title or Agreement

Figure 6: Sample Evaluation form for collecting data