Decision Support for Farmers in Africa

Analysis of the Roles and Competencies of Data Scientists in the Grain Industry

Martina Kotzé
University of the Free State Business School, Faculty of Economic and Management Sciences, South Africa

Yolandi Kruger
University of the Free State Business School, Faculty of Economic and Management Sciences, South Africa

Abstract

The use of data science as a decision-making tool in agriculture is becoming more evident. Data science principles and tools allow producers to access more consumable, accurate and timely information for making informed choices. Yet, the importance of farm data in effective decision making appears underrated in sub-Saharan Africa. To capitalise on 'big data', the sub-Saharan African grain industry needs competent data scientists to manage large data sets. To assist grain industry managers in selecting suitable candidates, an understanding of the data scientist's role and required competencies is essential. This research identifies the core competencies and explores the role of data science in the grain industry. The results reveal the different types of data scientists and indicate that their role should not be narrowly delineated.

Introduction

The world population will increase by 50 per cent in the next 30 years and sub-Saharan Africa will account for a third of this growth.[1] Most of the African population lives in the rural areas and relies on agriculture directly or indirectly for their livelihood,[2] while agribusiness and agro-industries account for over 30 per cent of national incomes.[3] Therefore, the agricultural sector, and specifically the role of agricultural transformation, is seen as the engine of growth[4] and the key driver of food security and poverty alleviation on the African continent.[5,6]

Since Africa and the South American continent have the most potential for new agricultural land development and both have sufficient water resources (outside the Sahara Desert), they have the potential to produce enough food for their own needs and to help meet the needs of the global population.[7] Although agricultural businesses want to undertake large-scale farming in Africa, most African agricultural production is still done by smallholders. A mixed-mode farming typology is also emerging in African agriculture and is evident in labour-intensive farm production activities where large-scale and smallholder agriculture co-exist through coordinated and collaborative activities based on both economies of scale and commercialisation.[8]

Recent economic growth has not spilt over to the agricultural sector with the exception of a few countries. For sub-Saharan Africa, current agricultural productivity is low and numerous initiatives have failed.[9] Furthermore, the agricultural sector has struggled to meet the growing demand of a wealthy consumer class.[10] In order to reduce food security problems and increase incomes, a transformation of the agricultural sector in sub-Saharan Africa is necessary.[11] This includes the creation of a dynamic agribusiness industry, able to produce efficiently and effectively.[12] Losses in African countries can be largely attributed to on-farm issues.[13] There is evidence that farm production can be increased through yield-increasing technologies and improved access to agricultural support services. Science and technological innovation have much to offer countries in Africa, including precision farming, the use of big data, GPS systems, drones and sensors to precisely detect yield, fertility and the moisture of soil. These may enable smallholders to increase their income levels, but require operation at a higher production function.[14,15]

Data Science and Agriculture

Decision making forms the basis of management in farming and, in order for farmers to increase the production of grain, effective decision making regarding their farms is essential.[16] External and internal sources are available to farmers to assist with effective decision making.[17] External sources refer mainly to other farmers, agricultural journals and state and private institutions, while internal sources refer to the farmer's own record system, or farm management information system. Farm data is essential for sound decision making by farmers and organisations providing support and advisory services to enhance agricultural production. Farm data is also indispensable for agricultural commercialisation as a strategic decision support resource at different levels of policy-making and management.[18] If provided with the correct input, feasible technology and relevant information, even small-scale farmers can transform traditional agriculture.[19] Up-to-date information, including the market prices of commodities and inputs, allows farmers to make more profitable decisions on production activities. In the private sector, farm data is essential for entities such as farm input suppliers and distributors, financial institutions, traders and other agribusiness service providers in order to assess opportunities and prospects in the agricultural sector. Farm data is also utilised by universities, colleges and research institutions for teaching purposes and research.[20]

Although certain role players who may be involved in farming indirectly, such as agricultural economists and scientists, may have access to large amounts of data, unfortunately it is not always tapped into by farmers and other role players for use in effective decision making at all levels.[21] A synthesis of ten national review papers on the status of farm data systems in sub-Saharan Africa indicates that the importance of farm data has tended to be underrated in the past two decades of participatory development.[22] Studies conducted in ten countries in Eastern, Western and Southern Africa found that access to reliable and timely farm information by farm producers, local communities and service providers is limited while also highlighting the limitations of the available data for decision support systems.[23] Currently, farm data systems in sub-Saharan Africa comprise of fragmented and unconnected multi-source systems that demonstrate momentous data gaps and a lack of co-ordination in data collection, analysis, utilisation and dissemination.[24] Furthermore, information systems in Africa often provide data that is outdated, of a poor quality and with the focus mainly on only a few commodities.[25] Early warning systems operating in Africa are mainly donor-founded and internationally managed, which has disadvantages for their sustainability.[26]

Data science is the management of large data sets from various sources to generate specific results which assist in informed decision making[27] and could be applied to improve both internal and external decision making.[28] The application of the principles and tools of data science in the grain industry will enable grain producers to access more consumable, accurate and timely information that can be used as a basis for making informed choices.[29] In recent years, several agricultural organisations have focused efforts on providing big data to farmers to assist them in making informed decisions based on real-time analyses of captured data. Projects have been started in developing nations which use large and small data sets and make use of complex approaches such as modelling disease diffusion or simple analyses enabled by recently available government data.[30] These projects can help improve quality of life in the developing world in several ways, including agriculture.[31] Although the Data Revolution, together with a high level of interest in data science, has arrived in Africa, the field is still nascent in this continent.[32] Unfortunately, the many opportunities for data to transform the developing world are not well-known. For the field to develop, it is therefore important to share experiences and cross-pollinate ideas on how data science can be applied successfully.

The Role of a Data Scientist

Data science concepts originated from a combination of different, existing disciplines[33,34,35] and have been implemented very successfully in other industries and sectors such as healthcare, retail, manufacturing and the public sector, as well as by Google.[36,37] To capitalise on big data, skilful data scientists would be required to manage and distribute the information. A data scientist is an information professional with the knowledge and skills to conduct sophisticated, systematic data analyses.[38,39,40] They are viewed as data engineers who use large quantities of data to gain insight into business and other complex systems.[41] Data scientists are generally analytically minded, and statistically and mathematically sophisticated.[42] They need to inform decision makers and information users of the value that can be derived from the application and innovative use of existing technology to organise, analyse and curate data.[43]

In the grain industry, data scientists work together with researchers and farmers to increase the production efficiency of staple crops like wheat, rice and corn.[44] Since data science is a relatively new discipline in the grain industry in Africa, it should preferably be introduced to farmers and the grain sector as a whole and its implementation and efficacy monitored.[45] Although the role of data scientists in other industries has been discussed to some extent in the literature,[46,47,48,49,50] empirical research regarding the role and competencies of a data scientist within the grain sector is limited. To capitalise on big data, organisations in the grain industry in sub-Saharan Africa would be required to recruit and appoint data scientists with the necessary skills and expertise to manage and distribute the information needed to improve decision making and increase productivity. In order to do so, the role and required competencies of a data scientist – specifically in the grain industry – need to be clarified.

Competencies and Competency Models

Competencies comprise general descriptions of the underlying attributes, knowledge and skills needed by individuals to ensure worthy performance in their specific occupation or job.[51] *Attributes* include the personal characteristics, traits, motives and values or ways of thinking that affect an individual's behaviour, while *knowledge* is the factual information that a person has and that is needed for a specific job.[52] A *skill* is an ability that has been acquired by training and education[53,54] enabling a person to consistently perform a complex task accurately, effectively and efficiently.[55] A skill is thus the demonstration of a particular talent.[56] Therefore, competencies are viewed as the set of behaviours instrumental in the delivery of desired results or specific outcomes.[57] Various authors collectively refer to related sets of knowledge, skills and abilities as a competency model and describe such a model as a selection of competencies required by a specific occupational group.[58,59,60] The development of a competency model for a specific occupational group allows an organisation to identify the behaviours that drive successful performance in a specific job or role and enables the organisation to deliver their technical expertise effectively.[61] Therefore, it is strongly advisable that a competency model should form the basis of recruitment and evaluation of potential candidates for specific positions.[62]

In light of the above, the objectives of this research were, firstly, to explore the role of data science in the grain industry by means of a qualitative study. Secondly, the research identifies and describes the core competencies of a data scientist in the grain industry and proposes a competency model that may form the basis for the recruitment and evaluation of potential data scientists in the grain industry in sub-Saharan Africa.

Research Methodology

A qualitative research method was applied. Qualitative research is exploratory and is useful when a new topic is addressed with a certain group of people.[63] The field of data science is fairly new and there is limited information on the topic. Furthermore, no examples of competency models for data scientists in the agricultural or grain industry are available. Therefore, the use of an exploratory study was most suitable. The essence of the role and competencies of a data scientist was identified and described by the participants of the study. Interviewing was used as the method of qualitative data collection.[64]

Sample

Only a limited number of people in a small category had the information that was sought for the purposes of the study. At the time of the study, according to our knowledge, no data scientists had been employed in the agricultural sector in sub-Saharan Africa and therefore no data scientists in this region could be included in the study. Since various important agricultural innovations originated from the US – including the use of data science in agriculture – the US was identified as a setting where the necessary information relating to this study could most likely be obtained. Non-probability judgement sampling was used[65] and 20 participants were interviewed from nine different organisations in the US who have a profound knowledge of data science and its role in the grain industry. The participants included individuals in the grain industry who work directly with data scientists in the same teams or are data scientists themselves.

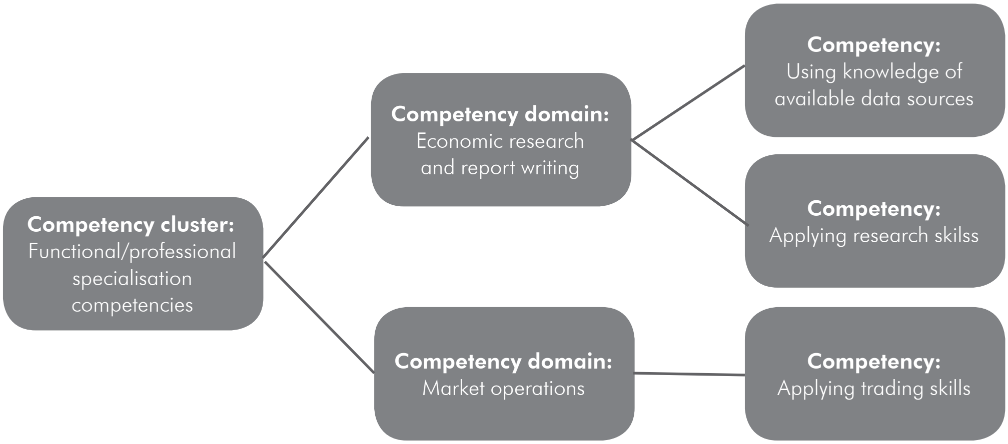
Data Collection Process

Semi-structured, face-to-face interviews were used for the purpose of the research. A list of semi-structured open-ended questions relating to the different objectives was prepared by the researcher.[66] Participants were requested to provide information on the role and competencies of data scientists in the grain industry. They were informed of the interviews beforehand by means of a briefing letter to ensure that they were well prepared and that they understood the purpose of the interview. All interviews were recorded with an audio recorder and transcribed afterwards. The participants were asked for their consent before the conversations were recorded.

Data Analysis

The data analysis strategy used in the study to identify the role and competencies of a data scientist in the grain industry included categorising strategies (thematic analysis and coding), connecting strategies (case studies and narrative analysis) and data display strategies.[67] Data was rearranged into different categories by means of coding[68] and tagging.[69] After coding and tagging the material, data with similar characteristics was placed in the same group or category. Once the data was categorised, the researcher applied connecting strategies.[70] Connecting strategies involve an attempt to understand the data in context by making use of different methods to identify relationships between the various elements in the text.[71] Data display gives direction on how to present the data and data coding assists in developing ideas while simultaneously drawing preliminary conclusions.[72] In order to develop the competency model, the steps proposed by Brits[73] were followed. Brits developed a conceptual framework for the development of a competency model that can be applied to any organisation. According to the definition offered by Brits,[74] a competency model consists of competencies, competency domains and competency clusters. A competency domain is a collective name for a group of similar competencies, while a competency cluster is a collective name for a set of domains (see Figure 1 below).

Figure 1: An example of the components of a competency model[75]



Competencies were identified from the data provided by the participants in terms of attributes, knowledge and skills and were assigned a descriptive label and a definition.[76] After listing the competencies, they were grouped into competency domains. The competency domains used were adapted from various competency libraries.[77,78,79,80] A competency library refers to 'a list of competencies from which to select when developing a competency model'.[81] In order to develop the competency clusters, links between competencies were made and the clusters identified by the Greater London Authority[82] were used as guidelines in this analysis.

The above process allowed the researcher to identify, label and prioritise the core competencies needed by a data scientist in the grain industry and to present them according to competency domains and clusters using the theoretical framework developed by Brits.

Results

The Role of Data Science in the Grain Industry

Role categories were identified based on the data generated from the interviews and ranked according to the frequency with which these were mentioned by the interviewees. A discussion of the identified categories follows, whereafter a summary of the role of data science in the grain industry is given, based on the identified categories.

Decision Making

Decision making, as part of the role of data science, was mentioned most often by the participants. Twelve out of the 20 participants indicated that data science drives broader, quicker and more efficient decision making. Farmers are perceived to be the data customers who make use of decision support tools, as developed by data scientists. One participant said that data science 'essentially helps you make decisions for the future year'.

One participant indicated that, for the past 20 years, large amounts of data have been collected by means of superficial analysis. The participant mentioned that there are companies that provide software programs a farmer or retailer can use for data analysis to provide cursory insights. However, there has been very little deep statistical analysis and almost no use of information with regard to the weather and how it impacts decisions. For this participant, a major advantage of data science lies in the use of all historical weather information collected over the last 30 to 50 years and making inferences on how the weather impacts crops. Furthermore, the participant feels that data science can be used to predict with relative certainty what is expected to happen throughout the next season. Another participant stated that most agricultural producers have 30 to 40 chances of generating a pay check in their entire career. Data science can be used to ensure a better probability of a good outcome for each one of those chances. This view is also supported by another participant, a grain producer.

Data Analysis

Data analysis was mentioned a significant number of times (seven) by the participants. One of the participants noted that farmers today have a lot of information about their fields, about their yields, about their machinery, but it is very hard for them to analyse that information and be able to decide what decision they can make that will impact profitability for their farm. Participants view data analysis as the step following data collection. Large volumes of data are already being collected and data science supports the analysis of the data. One participant explained the steps as follows: Analyse the data, separate the signal from the noise, so to speak, and get actual recommendations. A grain farmer and participant indicated that, before the application of data science in the grain industry, farmers were not able to analyse their planters' performance. Yet, data science allows them to do so.

Insights from Information

Five of the participants feel the role of data science involves forming insights from the information. According to one participant, 'True data science is statistics used to form insights out of the information'. Another participant stated that, once the information has been collected, managed and analysed and insights have been created, what happens with those insights is the concern of data scientists. A third participant explained the importance of weather and being able to calculate the effect of weather patterns on crop yields and 'start making inferences on how that weather impacts the crop, and predict with relative certainty what's going to happen throughout the next season'.

Problem Solving

Participants referred to the role of data science in agriculture as helping solve producers' problems by answering complex questions. One participant mentioned that data science involves solving producers' problems before they may even know that the problem exists. Data science is thus about 'having models in place or being able to quickly create a model to solve someone's problems'. Producers are faced with problems on a daily basis that require decisions on how to act on them. Another participant indicated that data science helps farmers to manage their operations more effectively in answering questions like 'when to plant, when to fertilise, when to irrigate, when to harvest and move equipment'.

Increased Profitability

By applying data science, producers can increase their profitability. Four of the participants mentioned this advantage. According to one participant, the role of data science in agriculture involves:

> … understanding how we take a big pool of data and start to create simple products that improve growers' operations efficiencies. Allow a grower to utilise, leverage and incorporate it in their business processes without a lot of complexity and it is kind of natural in their flow to increase productivity, yield, top line revenue or optimise cost.

Visualisation of Data

Visualising analysed data results in information that is easy to read, and also understandable. This category was mentioned by three participants in answering the question on the role of data science in agriculture. One participant stated, 'Everything is becoming so visual' and that 'people really want interactive visual data'. Farmers should be able to see in one picture what the effects of their decisions will be on the crop yield, for example, if they apply too little or too much fertilizer, how would that effect their yield and profitability?

Creating Value from Data

As part of its role in agriculture, data science must create something useful, such as a product. This is the view of three of the participants. One of the participants said that role players 'are starting to see value in their data, just like value in their crop'. Furthermore, he indicated that farmers are the users of data products using decision support tools that a farmer can use to make sense of the data they generate and to leverage it to make decisions.

Data Management

Data management was mentioned by two of the participants. From the interviews, it is evident that the need has arisen to manage large volumes of data collected on farms. According to one of the participants, 'precision agriculture', a concept which became popular in the US in the 1970s and 1980s, did not become what the industry envisaged at the time, because the computer systems needed to manage the data and answer complex questions which did not exist then. However, computer programs are now available for this purpose. Although a large amount of data is generated on a farm, the data is not compiled and analysed to bring solutions back to the farm.

Cleaning Data

Cleaning of data refers to excluding data that is inaccurate or irrelevant to the specific decision-making process. Two of the participants indicated that data cleaning should be regarded as part of the role of data science in agriculture. According to one participant, 'Data scientists are trying to make sure the data is as clean as possible and accurate'. Another participant indicated that only in the last few years has the industry become truly interested in the role of data scientists in cleaning data and gathering it in a way that allows it to be mined and analysed and for recommendations to be derived from it. For data scientists to build accurate decision-making systems they need to obtain all the details of what is happening in the field in order to generate accurate and clean results.

Data as a Commodity

One participant referred to producing data commodities as part of the role of data science in agriculture. They explained:

> People are starting to realise that data can be seen as a commodity – one which farmers have not had the privilege or ability to leverage in the past – but now they can with the application of data science. Data science delivers data products supporting decision making on all levels.

Different Types of Data Scientists

Two participants felt strongly that the role of a data scientist should be well defined and understood by the data scientist themselves for them to be successful within the organisation. Both these participants referred to the different types of data scientists. One participant stated:

> If you go to San Francisco, they will say data scientists … take the data collected and compiled by someone else, put the data into a database and create insights. That's a data scientist, but that's only one part of this big chain.

Another participant explained:

> I think it's important to define them because, in the absence of defining the different types of data scientists, I don't think you can be successful. It's going to be very difficult to find one person that has all of the expertise.

These participants went on to describe the different types of data scientists as they see it. Table 1 summarises the four types of data scientists, as identified by these two participants.

Table 1: Different types of data scientists, as described by the interview participants

Participant A	Participant B
Type 1: Data Science Collector	**Type 1: Equipment Expert**
One who collects data with precision and accuracy using the appropriate equipment and uploads various meta-data into the cloud. **He collects data relating to the weather, the crop being planted, the seeding rate, genetics that were used, etc.**	An individual who understands all the tools and measuring devices being used to collect accurate data, for example, different yield monitors and the fact that each model provides different outputs.
Type 2: Software Systems Engineer	**Type 2: Field-based Research Scientist**
Someone who receives the meta-data from the cloud and decides how to store it and develop an interface to access the data so that the next user can analyse it.	A scientist who conducts field research and manages the interface with the platform user. Field research includes measuring statistical differences in terms of performance.
Type 3: Data Analyst	**Type 3: Modeller**
The analyst analyses data, conducts data modelling and derives insights from the data.	The modeller analyses data, conducts data modelling and derives insights from the data.
Type 4: Data Science Applier/Agronomist	**Type 4: Agronomic Scientist**
This individual applies the insights received from the data analyst and converts these into actual recommendations which are conveyed to producers. **He/she knows whether the recommendations are practical for farming and grain production conditions.**	The agronomic scientist understands the interaction between soil and crop physiology, and how to apply insights and make useful recommendations.

Source: Authors

From Table 1 it can be seen that there is duplication in the different types of data scientists described by the participants. Previous literature divides data scientists into four types, namely, data business people, data creatives, data developers and data researchers.[83] The explanations from the participants support the theoretical framework of Harris et al.[84] and the results indicated that the types of data scientists identified during the empirical research can be categorised according to this framework. Table 2 displays the new categories as derived and adapted from the literature and the empirical results of this study.

Table 2: Different types of data scientists – combined categories

Combined Categories	Participant A	Participant B
Data Science Coordinator	**Type 1 and 4**	**Type 1 and 4**
The data science coordinator is concerned with their organisation and to what effect data projects can be initiated and managed. They see themselves as entrepreneurs and leaders with technical data science skills as well as a profound knowledge of the domain they work in. They know all the role players – who can contribute, who needs support and who will utilise the data products – as well as the relationships between these role players.		
Data Analyst	**Type 3**	**Type 3**
Data analysts extract and integrate data, perform advanced analysis, create visualisations, conduct interpretations and build tools to make the analysis scalable and applicable to other users. Data analysts usually have academic experience with undergraduate degrees in the field of economics or statistics. They can rapidly transform data into something of value. Harris et al. describe this group as representative of the broadest view of data scientists.		
Data Developer	**Type 1 and 2**	**Type 1**
Data developers usually have a computer science or computer engineering degree and are concerned with the technicality of data, that is, how to obtain, store and learn from it. Their daily activities include coding and machine learning. They should be knowledgeable of the different types of equipment used to collect and measure data with precision and accuracy.		
Data Researcher	**Type 4**	**Type 2**
Data researchers have a sound academic research background in the agricultural sciences, grain industry or a related domain. They apply insights received from data analysts, converting these into useful recommendations that are relevant to the specific domain.		

Source: Authors

Summary of the Role of Data Science in the Grain Industry

From the information in Table 2, it is evident that the participants regard the role of data science as follows:

- facilitating effective decision making and problem solving in agriculture, specifically the grain industry by means of computerised data management;
- cleaning data, data analysis and the visualisation of analysed data in order to derive profitable insights;
- data science creates value from data by turning information from the farm operation into decisions that lead to increased profitability; and
- data scientists can play different roles, including being data science coordinators, data analysts, data developers and data researchers.

The Core Competencies of a Data Scientist in the Grain Industry

Table 3 presents the categorised core competencies of a data scientist in agriculture based on the data gathered from the interviews. The competencies are listed according to knowledge, skills and attributes. The table includes the identified category, a description of the category and the frequency it was mentioned by the participants. The list is sorted according to the frequency mentioned, which underlines the priority assigned to the specific competencies. A discussion and summary of the knowledge, skills and attributes is given after the table.

Table 3: Core competencies of a data scientist in agriculture in terms of knowledge, skills and attributes

Identified category	Description	Frequency mentioned
Knowledge		
Knowledge and experience of the grain industry	• Knowledge of the grain industry including farming, plant breeding, agronomy, soil science or some aspect of these • Real farming experience is highly valuable.	9
Knowledge at a post-graduate level	• Masters or PhD-level qualification in physics, computer science, data science, statistics, other related sciences or agronomy	7
Agronomic science	• Understands interaction between soils and crop physiology, and how to apply insights from this and make recommendations within the grain industry	4
Mathematics	• Knowledge on how to apply a mathematical strategy to solve a practical problem	4
Data collection	• Knowledge of data collection equipment and format of data collected by the device, particularly how to upload the data from the device to a cloud or other means to access it • Geo-spatial information system knowledge and its effect on the grain industry	4
Software engineering systems	• Bachelor to PhD in computer science or mathematics • Expertise in computer languages such as R or Python	2
Statistics	• Sound knowledge of statistics	2
Data analysis and modelling	• Knowledge of the relationship between data and outcomes • Knowledge of how to rewrite complex problems in a mathematical model	2
Field-based research science	• Knowledge of how to conduct the field research component of data science and manage the interface with the user of the platform • Knowledge of how to conduct field research to measure statistical differences in terms of performance	1

Equipment expert	• Understands all the tools and measuring devices used to collect accurate data including yield monitors, GIS systems, precision agriculture tools	1
Machine learning	• Knowledge of machine learning algorithms and how to use each one for solving different problems	1
	Skills	
Communication skills	• Able to effectively transfer thoughts and express ideas verbally in individual or group situations; communication to the grain sector as a whole or to specific farmers	12
Data analysis	• Statistical skills • Able to apply analytics on available data in order to gain insights from it	7
Software engineering systems	• Computer programming skills • Data processing skills • Coding skills including experience in translating models into code • Data ingestion skills • Able to develop the right interface to allow third group analysis	6
Achieving results and satisfying customer expectations	• The ability to take direct action in order to attain or exceed objectives • Develops alternatives when certain actions have not led to a desired result • Uses knowledge and expertise in data science and applies it to practical situations • Actively seeks the best way to achieve goals	5
Quantitative skills	• Quantitative mind-set, ability to code and use different statistical software • Skills for manipulating data	3
Data modelling	• Able to conduct data modelling and extract value out of this • Able to link different processes, place correct weighting on each one and connect these with the help of the agronomist so that the output is accurate	3
Teamwork	• Cooperative and works well in a team to find solutions which generally benefit all involved parties	2
Visualisation of data	• Able to present data visually in a way that is easy to see, read and understand	1
Writing skills	• Able to transfer thoughts and ideas onto paper and create reports	1
Multitasking	• Ability to work on multiple projects at the same time	1
Field-based research skills	• Ability to understand trials and trial results; trials in the grain industry involve planting small areas of land – testing a set of predefined elements and its effect on crop yield and profitability	1

Attributes		
Creativity	• Ability to come up with original and innovative ideas and solutions and to adopt points of view outside the usual parameters	11
Attention to detail	• Thoroughness in accomplishing a task with concern for all areas involved, no matter how small • Monitors and checks work or information and plans and organises time and resources efficiently	7
Collaboration	• Fosters cooperation and teamwork while participating in a group and working toward solutions which benefit all involved parties	6
Continuous learning	• Demonstrates eagerness to acquire necessary technical knowledge, skills and judgement to accomplish a task effectively • Desire to acquire knowledge and skills necessary to perform job more effectively • Ability to absorb new information readily and put it into practice effectively • Recognises mistakes and attempts to correct or prevent them	5
Accuracy	• Consistently delivers work of a high quality that is precise and meets standards, procedures, rules, regulations and expectations • Recognises and separates quality data from insignificant data	3
Receptivity	• Open to dialogue and a good listener • Absorbs and understands important verbal and non-verbal information and asks further questions when necessary • Ability to understand their audience	3
Problem solving	• Generates creative approaches to problems and opportunities	3
Passionate	• Engaging and enthusiastic	2
Questions traditional methods	• Considers non-traditional farming practices	1
Vision	• Ability to see beyond daily tasks and explore ideas for the future; regards the facts objectively and sees them in the broader context and over a longer term • Understands all the different factors a farmer has to consider in producing a crop and how they are connected • Connects variables with outcomes	1
Investigative	• Asks the right questions on technical and functional matters	1
Tenacious	• Shows persistence in finding a solution despite hurdles, as most problems are very complex	1
Agile and flexible	• Able to use a variety of devices from different manufacturers	1

Source: Authors

Knowledge

Overall, *knowledge and experience of the grain industry* is the most often mentioned in this category. Participants strongly felt that data scientists need knowledge of farming, the grain industry, plant breeding, agronomy and science or some aspects of these disciplines. This knowledge could be obtained through personal experience in the domain, for example, if someone grew up or lived on a farm for some time. The data scientist should be familiar with the lifecycle of different crops, know what elements have a significant impact on crop yields and how to produce at an optimum level. The importance of knowledge at a *postgraduate level* in *physics, computer science, data science, statistics or other sciences* is the second most often mentioned knowledge competency. Knowledge of *agronomy* was frequently cited, including an understanding of the interactions between soil and crop physiology. How to apply insights and make recommendations was also referred to.

Furthermore, knowledge of the application of mathematical strategies to problem-solving, data collection (including from geo-spatial information systems) and knowledge of software engineering systems (including computer languages like R or Python and computer science) are valuable. Some participants include knowledge related to *statistics, data analysis and modelling, field research, equipment and tools* used to collect data, as well as an understanding of *machine learning algorithms* and how to apply each one for different problems.

Skills

Most participants regarded *communication skills* as a core competency. Data scientists should be able to communicate clearly and articulately when speaking with an individual or before a group, ensuring that others fully comprehend the intended message. Communication refers specifically to grain farmers and other role players in the industry. A good candidate is cooperative and works collaboratively in a team.

Secondly, data scientists need to have *data analysis* skills such as the ability to apply analytics on available data in order to gain insights from it, including *statistical skills*. For data scientists to properly analyse data, they need to use *software engineering systems*, ingest the data and develop the right interface. They should have adequate *computer programming* and *data processing skills*.

Other skills that were mentioned include *achieving results and satisfying customer expectations*, the ability to take action in order to attain or exceed objectives, qu*antitative skills* related to manipulating data and *data modelling* to extract value. Skills each mentioned by a single participant are *visualisation of data, writing skills, multitasking* and *field-based research,* which refers to the ability to understand trials and trial results. The trials involve planting small areas of land – testing a set of pre-defined elements and its effect on crop yield and profitability. The data scientist should be familiar with the elements that are tested, as well as the desired results in the context of grain production.

Personal Attributes

Creativity seems to be an important attribute and was mentioned by 11 of the participants. It refers to an individual's ability to generate original ideas. *Attention to detail* was mentioned by seven participants and refers to thoroughness in accomplishing a task and concern for all areas involved, no matter how small. Six participants identified *cooperation and teamwork* while participating in a group and working toward solutions which benefit all involved parties as important.

Following these top three attributes is an *eagerness to acquire the necessary technical knowledge, skills and judgement* to accomplish a task effectively. This was mentioned by five participants. The data scientist must have a desire to acquire knowledge and skills necessary to perform a task more effectively, recognise mistakes and attempt to correct or prevent them.

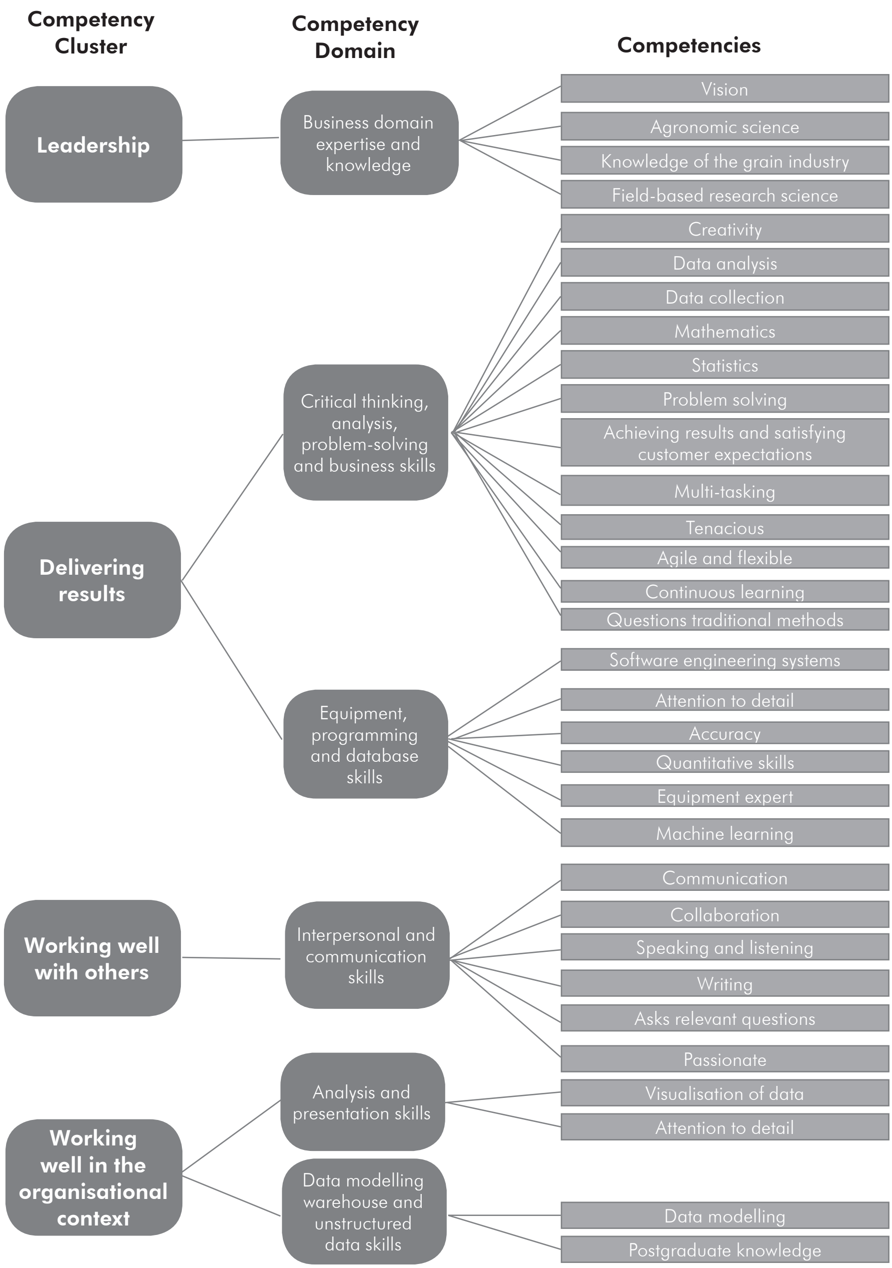
Attributes pointed out by three participants are *accuracy,* which requires the scientist to consistently deliver high quality accurate work that meets expected standards, being a *good listener* in order to understand their audience and *problem-solving* or generating creative approaches to address both problems and opportunities.

Two of the participants regard being *passionate* and engaging as required attributes, while the least frequently mentioned attributes (once each) include *questioning traditional methods,* having *vision, asking relevant questions* on technical and functional matters and being *tenacious, agile and flexible.* In the following section, the core competencies of a data scientist in agriculture are classified in terms of competency domains, clusters and individual competencies.

The Proposed Competency Model

In order to compile a competency model for a data-scientist in the grain industry, the above competencies were classified according to competency domains, clusters and individual competencies.[85,86,87,88,89] Figure 2 below provides a graphical presentation of the competency model and displays the core competencies of a data scientist classified in terms of competency clusters, domains and competencies.

Figure 2: A graphical representation of the core competencies for a data scientist in agriculture in terms of competency clusters, domains and individual competencies



Source: Authors

Discussion of the Results

The empirical data from the interviews shows that the role of the data scientist in the grain industry includes mostly facilitating effective decision making and problem solving in agriculture by means of computerised data management, the cleaning of data, data analysis and the visualisation of analysed data in order to derive profitable insights. These results are in line with previous literature that indicates that data scientists contribute to the collection, cleaning, transformation, analysis, visualisation and curation of large, heterogeneous data sets.[90]

However, both the literature[91] and the empirical results from this study show that there may be no universal role definition for data scientists in all organisations in the grain industry. The descriptions gained from both the literature and the empirical data were combined to classify the role of data scientists into four types, namely, data science coordinators, data analysts, data developers and data researchers. The most appropriate role for the data scientist would need to be identified depending on the context within which the organisation operates, as well as the objectives of the organisation. Contexts can differ in terms of the education levels of the various role players, as well as technology available to role players to access or distribute the data. In Africa, not only have the education levels of rural people increased substantially, but the change in communications technology has brought information and knowledge much closer to small farm households.[92] The functionality of mobile phones is likely to grow dramatically in the coming decades for many forms of information provision, such as decision support tools. Mobile technology may facilitate the availability and improvement of an increasingly complex arrangement of data for analysis and prediction among businesses and communities. Therefore, in Africa, the form of big data is somewhat different compared to other parts of the world as most of the external data is generated from mobile devices.[93]

Furthermore, in terms of competencies, it follows that the data scientist needs a solid foundation in the specific domain knowledge (grain industry), together with a Masters or PhD-level qualification in physics, computer science, data science, statistics or other related sciences or agronomy. Domain knowledge of grain production, plant breeding, agronomy, soil sciences and/or real farming experience also seems necessary and highly valuable. In terms of skills, proficiency in data analysis and statistics and effective communication are essential. Candidates should be able to apply analytics on available data in order to gain insights from it.

The most important attribute of an effective data scientist in the grain industry, according to the participants, is being creative; that is, having the ability to come up with original and innovative ideas and find or design solutions. Furthermore, attention to detail, thoroughness in accomplishing tasks with a concern for all areas involved, and the ability to foster co-operation and teamwork while participating in a group and working toward solutions which benefit all concerned parties, are essential. It was also evident from the results that data scientists need to demonstrate an openness towards continuous learning and an eagerness to acquire the necessary technical knowledge, skills and judgment in order to perform their roles more effectively. An individual who is not afraid of setting challenging goals and pursuing new avenues to achieve these goals is best suited to the role.

The competencies identified in this study cannot be generalised to the grain industry in various contexts since the study only included data scientists in the US. However, it may be used as guidelines for the development of job and person profiles for data scientists in sub-Saharan Africa.

Conclusion

Producers and agribusinesses are recognising the value derived from data science as a decision-making tool and it is becoming more widely used. The results from the interviews show that the use of data science in the grain industry is a growing trend and that its importance should not be underestimated. Data scientists support producers by supplying more timely information in a simple format to help them make quicker and well-informed decisions. There are different types of data scientists and the role of the data scientist should not be narrowly delineated. The differentiation should not be based on the individual's breadth of knowledge, but rather on their depth of knowledge in a specific area and in relation to other relevant fields, as well as their preferred methods of addressing data science problems.[94]

Certain limitations of the study were identified. Participants included in the study all reside in the US. Therefore, it is recommended that future studies be conducted in organisations in the grain industry in developing countries where data scientists are being employed. The competency model also needs to be validated and tested by interviewing experts and preferably also human resource managers in the grain industry, to add value to this research.

Competencies in the form of behaviours, skills and attributes, as well as a generic competency model, have been proposed based on the empirical results. However, it is important for organisations in sub-Saharan Africa intending to appoint a data scientist to begin by clarifying the role(s) and competencies required, based on the different role definitions. Such organisations should also clearly define the context in which the data scientist is required to function before recruiting. This would be beneficial to the job holder, the organisation and the other role players involved.

Notes and References

1 World Population Review. 2014. African Population 2014. Available at http://worldpopulationreview.com/continents/africa-population/, [Accessed 13 August 2015].

2 Diao, X., Hazell, P., Resnick, D. and Thurlow, J., 2007. The role of agriculture in development: Implications for Sub-Saharan Africa. Research Report 153, International Food Policy Research Institute, USA: Washington DC.

3 Connolly, A.J., 2014. A Glimpse into the Future: A Lens through which to Consider 'Africa's Rising'. *International Food and Agribusiness Management Review*, 17, Special Issue B, pp. 9–18.

4 Chauvin, N.D., Mulangu, F. and Porto, G., 2012. Food production and consumption trends in Sub-Saharan Africa: Prospects for the transformation of the agricultural sector. Working Paper 2012-011, UN Development Programme.

5 World Bank, 2013. Growing Africa: Unlocking the potential of agribusiness. AFTFP/AFTAI. Washington DC: World Bank.

6 NEPAD, 2013. Report on the World Economic Forum. Cape Town, 8–10 May. Available at http://www3.weforum.org/docs/AF13/WEF_AF13_Report.pdf, [Accessed 17 February 2015].

7 Connolly, 2014.

8 Van Rooyen, C.J., 2014. Towards 2015. Trends and Scenarios for African Agribusiness. *International Food and Agribusiness Management Review*, 17, Special Issue B, pp. 19–39.

9 Chauvin et al., 2012.

10 Nkomo, M., Shelman, M., and Connolly, J., 2014. African agribusiness on the move. *International Food and Agribusiness Management Review*, 17, Special Issue B, pp. 5–7.

11 Chauvin et al., 2012.

12 Nkomo et al., 2014, p. 5.

13 Connolly, 2014.

14 Ibid.

15 Van Rooyen, 2014.

16 Shiya, M.A., 2011. Implementation of the South African excellence model (entry level) to measure and improve performance of SMEs in agriculture and related businesses. Bloemfontein: Central University of Technology.

17 Ibid.

18 Minae, S., Baker, D. and Dixon, J., n.d. Status of farm data systems and farmer decision support in Sub-Saharan Africa. Available at http://ftp.fao.org/upload/Agrippa/628_en.doc, [Accessed 10 May 2016].

19 Ozowa, V.N., 1995. Information needs of small scale farmers in Africa: The Nigerian Example. *Quarterly Bulletin of the International Association of Agricultural Information Specialists, IAALD/CABI,* 40(1).

20 Minae et al., n.d.

21 Kruger, Y., 2015. A competency model for data scientists in GrainSA. Unpublished field study, Magister in Business Administration, UFS Business School, University of the Free State, Bloemfontein, South Africa.

22 Minae et al., n.d.

23 Ibid.

24 Ibid.

25 Antonaci, L., Demeke, M. and Vezzani, A., 2015. The challenges of managing agricultural price and production risks in sub-Saharan Africa. ESA Working Paper No. 14-09. Rome, FAO.

26 Ibid.

27 Umachandran, K., 2013. Data Science for Agriculture. Available at https://www.researchgate.net/publication/262877679, [Accessed 15 May 2015].

28 Monsanto, 2014. Launch of Open Ag Data Alliance supports farmer privacy and control of their data, enables adoption of data science-powered agriculture. *Business Wire*, 11 March 2014. Available at http://news.monsanto.com/press-release/research-and-development/launch-open-ag-data-alliance-supports-farmer-privacy-and-cont, [Accessed 12 April 2015].

29 Monsanto, 2014. The Role of Data Science in Agriculture. Available at http://news.monsanto.com/image/infographic/role-data-science-agriculture, [Accessed 30 September 2014].

30 Korte, T., 2014. How data and analytics can help the developing world. Center for Data Innovation. Available at https//datainnovation.org/2014/07/how-data-and-analytics-can-help-the-developing-world, [Accessed 9 May 2016].

31 Quinn, J., 2015. A new generation of data science innovators in Africa. United Nations Global Pulse. Available at http://unglobalpulse.org/breeding-data-science-innovators-Africa, [Accessed 9 May 2016].

32 Ibid.

33 Stanton J., Palmer, C.L., Blake, C. and Allard, S., 2012. Interdisciplinary Data Science Education. *Special Issues in Data Management*, 1110, pp. 97–113.

34 Loukides, M., 2012. What is Data Science? The future belongs to the companies and people that turn into products. Sebastopol, CA: O'Reilly Media Incorporated.

35 Harris, H.D., Murphy, S.P. and Vaisman, M., 2013. Analysing the Analysers – An Introspective Survey of Data Scientists and Their Work. Sebastopol, CA: O'Reilly Media Incorporated.

36 Loukides, 2012.

37 Manyika, J., Chuis, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H., 2011. Big data: The next frontier for innovation, competition, and productivity. Toronto: McKinsey Global Institute.

38 Hosack B., Power, D. and Sagers, G., 2014. Designing Data Science Graduate Programs: A Case for Applied Doctorates in IS. MWAIS Proceedings. Paper 16. Association for Information Systems Electronic Library. Available at http://aisel.aisnet.org/mwais2014/16, [Accessed 21 April 2015].

39 Power D., 2015. What universities offer master's degrees in analytics and data science? Available at http://dssresources.com/faq/index.php?action=artikel&cat=id=250&artlang=en, [Accessed 1 May 2015].

40 Stanton et al., 2012.

41 Hillion, S., 2011. What is a data scientist? Available at http://www.forbes.com/sites/danwoods/2011/10/11/emc-greenplums-steven-hillion-on-what-is-a-data-scientist, [Accessed 25 August 2014].

42 Ibid.

43 Stanton et al., 2012, p. 108.

44 Davidson, A., 2014. Meet a data scientist who's helped revolutionize agriculture. Forbes Science Business: Research, innovation and policy. Available at http//www.forbes.com/sites/sciencebiz/2014/11/20/meet-a-data-sientist-whos-helped-revolutionze-agriculture, [Accessed 9 May 2016].

45 Kruger, 2015.

46 Davenport, T.H. and Patil, D.J., 2012. Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 19(10), pp. 70–76.

47 Harris et al., 2013.

48 Loukides, 2012.

49 Sanders, M., 2013. Data Scientist Core Skills. Available at http://www.datasciencecentral.com/profiles/blogs/data-scientist-core-skills?xg_source=activity, [Accessed 12 October 2014].

50 Stanton et al., 2012.

51 Coetzee, M. and Schreuder, D., 2013. Personnel Psychology: An Applied Perspective. Cape Town: Oxford University Press.

52 Akkermans, J., Brenninkmeijer, V., Huibers, M. and Blonk, R.W.B., 2012. Competencies for the contemporary career: Development and preliminary validation of the Career Competencies Questionnaire. *Journal of Career Development*, 40(3), pp. 245–267.

53 Brits, D.W., 2012. Building and validating a competency model delivered by a corporate university. Johannesburg: University of Johannesburg.

54 Croucamp, A., 2013. Developing a Competency model for Head of Departments at Tertiary Education Institutions in South Africa. Cape Town: University of Cape Town.

55 Murphy, K.R., 2010. Individual Difference that influence performance and effectiveness. What should we assess? In Scott, J.C. and Reynolds, D.H. (eds.) Handbook of Workplace Assessment. Los Angeles: Wiley.

56 Mirabile, R., 1997. Everything you wanted to know about Competency Modelling. *Training and Development,* 51(8), pp. 73–77.

57 Brits, 2012.

58 Ibid.

59 Bartram, D., 2012. The SHL Universal Competency Framework. SHL White Paper 2012.

60 Campion, M.A., Fink, A.A., Ruggeberg, B.J., Carr, L., Phillips, G.M. and Odman, R.B., 2011. Doing competencies well: Best practices in competency modelling. *Personnel Psychology,* 64, pp. 225–262.

61 CIPD, 2014. Competence and Competency Frameworks. Available at http://www.cipd.co.uk/hr-resources/factsheets/competence-comptency-frameworks.aspx, [Accessed 19 October 2014].

62 Campion et al., 2011.

63 Creswell, J.W., 2003. Research Design Qualitative, Quantitative, and Mixed Methods. Lincoln, TX: Sage.

64 Schurink, W. and Schurink, E., 2012. Workshop in Qualitative Research (course notes). Presented at the University of the Free State, Bloemfontein, South Africa, 29–31 May 2012.

65 Sekaran, U. and Bougie, R., 2013. Research Methods for Business: A Skill-Building Approach (6th edition). West Sussex: John Wiley & Sons.

66 Ibid.

67 Maxwell, J.A., 2009. Designing a Qualitative Study. In Bickman, L. and Rog, D. (eds.) The Sage Handbook of Applied Social Research Methods. Los Angeles: Sage, pp. 214–253.

68 Strauss, A., 1987. Qualitative Analysis for Social Scientists. Cambridge: Cambridge University Press, p. 29.

69 Baptiste, I., 2001. Qualitative Data Analysis: Common Phases, Strategic Differences 2. QDA: Four Phases, Numerous Strategies. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 2(3), p. 26.

70 Ibid.

71 Maxwell, 2009.

72 Sekaran and Bougie, 2013.

73 Brits, 2012.

74 Ibid.

75 Ibid.

76 Campion et al., 2011.

77 Harvard University, 2015. Harvard University Competency Dictionary. Available at https://www.campusservices.harvard.edu/system/files/documents/1865/harvard_competency_dictionary_complete.pdf, [Accessed 1 May 2015].

78 HRSG, 2015. Accelerate competency-based management in your organization with competency dictionaries. Available at http://www.hrsg.ca/wp-content/uploads/2014/08/HRSG-Dictionary-Overview-List.pdf, [Accessed 1 May 2015].

79 TMA, 2012. TMA Competency Library. Available at http://www.tmamethode.nl/downloads/tmaCompetencymodel1-120eng.pdf ,[Accessed 1 May 2015].

80 WSU, 2015. Performance Planning and Development Program. Competency Library. Detroit: Wayne State University.

81 Campion et al., 2011, p. 245.

82 Greater London Authority, 2014. Competency Framework. Guide for Managers and Staff. Available at https://www.london.gov.uk/sites/default/files/Competency-framework-guidelines.pdf, [Accessed 19 October 2014].

83 Harris et al., 2013.

84 Ibid.

85 Harvard University, 2015.

86 HRSG, 2015.

87 TMA, 2012.

88 WSU, 2015.

89 Greater London Authority, 2014.

90 Stanton et al., 2012, p. 97.

91 Harris et al., 2013.

92 Garrity, D., Dixon, J. and Boffa, J.M., 2012. Understanding African farming systems: Science and Policy Implications. Paper prepared for the conference Food Security in Africa: Bridging Research and Practice. Sydney, Australia, 29–30 November 2012.

93 Ibid.

94 Harris et al., 2013.

