



Designing a Big Data Analytic Tool for Predicting Girl Child Learner Drop Out in the Eastern Cape Province of South Africa

Nosipho Mavuso¹ , Nobert Jere¹ , and Nelly Sharpley² 

¹ Walter Sisulu University, Nelson Mandela Drive, Mthatha, Eastern Cape, South Africa
nmavuso@wsu.ac.za

² University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa

Abstract. The benefit of education in the socio-political development of a country cannot be underestimated. Research has shown a strong correlation between education and socio-economic factors. In South Africa, approximately, 60% of young South Africans drop out of school, without finishing Grade 12 and obtaining their matric certificate. This can be attributed to various factors ranked by the study findings such as pregnancy, drug and substance abuse, socio-economic factors, and rape. It is thus important to address these challenges early. In essence, it is crucial to determine the factors that expose the girl-child to the risks of dropping out of school. In this paper, we present the design and development of a tool to predict the likelihood that a learner would drop out of school. The tool makes use of the risk factors determined through a qualitative approach, through interviews, focus group discussions and workshops with principals, teachers, and the School Governing Body. Pregnancy was highly ranked as the main contributing factor for learners dropout rate. Using the tool, we predicted accurately the likelihood of a learner being a dropout, by providing 3 levels of ranking, high risk, medium risk, and low risk. In addition, the tool also provided insights into the relationship between the student's living arrangements, the distance they travel to school and the financial standpoint of the parents. Developing analytic tool using data from remote areas enable smart city planners to consider disadvantaged communities. Smart schools may be difficult to achieve if dropout issues are not addressed.

Keywords: Big Data Analytics · Drop Out · Prediction · Girl Child

1 Introduction

Learner dropout has been a topic of interest in South Africa in the past decade. This is primarily due to the increase in both global and national reports of students who drop-out of high school. The challenge of school dropout is not trivial due to the complexity and number of factors involved, many of which have been well documented in the literature.

Dropout harms the life of learners and has increased the pressure on economic development in many countries [1]. Furthermore, reports in the literature have pointed to a possible relationship between learner dropout and criminal behaviour among adolescents. The work in [2] reported a correlation between learner dropout and child abuse.

The current challenges facing learners and communities such as poverty, family issues, health challenges, inadequate infrastructure at the schools and the community environment have led to high dropout numbers. As a result, research on dropouts has become a major interest to many stakeholders with government departments trying to reduce and avoid dropout [3]. Several dropout monitoring strategies, ideas and techniques have been proposed. As a result, some monitoring solutions have produced positive results, such as the various features of Learning Management Systems, that provide the identification of students at risk. One of the main challenges has been that dropout is only identified or acted on when a learner is already out of school [4]. In addressing the challenges, technologies and software developed have proposed and developed systems and applications that assist in dropout management [5]. Some developers have applied artificial intelligence, machine learning and big data analytic techniques to attempt to address dropout.

Using big data, a pattern of declining performance over time is a good predictor of student drop-out, and having dependents, being married, or serving in the military lowers the risk of dropping out. One of the technologies used is data mining, whose primary goal is to discover patterns, profiles, and trends through data analysis using advanced data analysis techniques and pattern recognition technologies [6]. The risk of dropping out was higher among older students, female students, and students with prior college education or transfer credits. Tools for data mining are used to discover data, gain a deeper comprehension of it, and predict future behaviour [7]. In a similar vein, [8] utilised data mining to locate educational data on dropouts and identified additional issues, such as privacy concerns, that must be considered in the studies.

According to [9] a list of data mining methods, which were divided into artificial intelligence and statistical method methods, approximately, 79% (22 of 28 studies) used decision tree classifiers. This method is utilised, as stated by [10] because of its adaptability when processing numerical and categorical data, its monotonous transformations of explanatory variables, and the ease with which results can be interpreted. In addition, it provides higher accuracy rates. The ID3 (Decision tree classifier) algorithm is more sensitive than other algorithms when it comes to classifying data from student history registers. Since these data mining approaches are regarded as powerful tools for solving classification problems and are utilised frequently for their simplicity and ease of understanding, neural network classifiers and support vector machines hold the second highest frequency of use [11]. Linear Regression and Logistic Regression are popular methods for classifying data based on its characteristics and are adaptable to the use of categorical and continuous predictor variables.

This paper presents a big data analytic tool that was developed based on the girl-child dropout factors as pointed out by participants from OR Tambo District Municipality in Eastern Cape Province in South Africa. The dropout factors as given by the engaged learners were considered for the development of the dropout analytic tool. There are main factors that are attributed to learner dropout. The paper presents a big data analytic tool that was developed based on the factors that were collected from school-going girls between Grades 8–12. The understanding of different factors affecting learners' dropout provides an informative environment that assists in future planning.

In the following section, we present the literature review, Sect. 2 showcases the description of the case, and Sect. 3 is the Girl Child School Dropout Model. Section 4 is the methodology. Section 5, the analysis of dropout factors subsequently follows. Lastly, the conclusion and references are given in Sects. 6 and 7.

1.1 What Are the Factors Affecting Dropout?

The phenomenon of school dropout is perceived as a process that is caused by a variety of factors. Most studies have indicated several factors related to learner dropouts globally [12]. The following are the most cited factors generally leading to learner dropouts and before the Covid-19 pandemic crisis; pregnancy, parents' educational status, age at entry to first grade, substance abuse, bullying, family responsibility, school distance, teachers' attitudes, poverty, poor academic performance, failure to cope with school, a lack of social skills to cope with life's challenges, early employment, a lack of parental care and role models, child-headed families, media influence, poor payment of teachers, child labour, early marriage for young girls, and a lack of school/personal effects, to mention a few [13–16]. These factors have a huge bearing on how disadvantaged communities could be developed and as Africa moves towards a smart environment, there is no room to ignore these factors.

These factors seem to differ in ranking per socioeconomic status, country to country and between low-income countries and high-economic countries. Over and above the noted ongoing crisis of learner dropouts worldwide, the Covid-19 pandemic added much to the learner dropout crisis globally, even so, the countries on the periphery were mostly affected due to a lack of resources to confirm the online learning spaces and ensure continuity in learning [17]. A joint UNESCO, UNICEF and WORLD BANK report in 2021 presented the global disruption to education caused by the Covid-19 pandemic as constituting the worst education crisis on record. The magnitude of the shock is still not fully understood, but emerging evidence is deeply concerning about the state of the global education crisis. Schools closed in 19 out of every 20 countries around the world, for a median of 17 weeks [18]. Inadequate facilities, such as a lack of computers, and internet facilities, were the major factors that limited learner engagement. It was also observed that Covid-19 disrupted educational activities and reduced educational opportunities for disadvantaged people [19]. Covid-19 displaced students and teachers and created multiple barriers to teaching and learning. These effects were felt by educational institutions, educators, students, parents and other stakeholders in education [20]. Globally, studies on learner dropouts at the time of the Covid-19 pandemic relate learner dropout to consequences like learning poverty, economic impact, and learners older than the current grade [19]. The variations in the gaps between those who have and have not need to be considered in driving smart schools.

1.2 Technologies/Innovation Challenges at Schools

Technology has been assimilated well in various industries, including service industries such as retail, transport and education, thereby giving prominence to the development of the Fourth Industrial Revolution [21]. However, in education, digital transformation has been lagging behind although technology has been one of the major innovations

which can potentially model the educational landscape. The education landscape in second world countries and marginalised communities has been faced with resistance in transformation into digitalisation due to a lack of skills, high cost of infrastructure and restricted application [22]. Regardless, the pandemic has catapulted the transformation of the education sector to digital learning and has forced even marginalised schools to reconsider their position in the digital transformation [22]. Digital transformation in the field of education is focused primarily on the use of computers as well as advancing the learning environment into sharing of materials and data analytics that assist in comprehending the needs of both educators and learners [23]. The incorporation of technology into traditional teaching and learning is, however, challenged by external and internal barriers to the teacher's implementation of technology. External barriers are first-order constraints. These include access constraints, teacher training and support.

Access constraints are challenges to do with an inadequate supply of connectivity, including the equipment. Without adequate computer hardware, software, and internet connectivity, it is impossible to implement technologies in schools. The provision of computers is the most basic stage in the introduction of technology in educational institutions. Many schools are on a drive to provide computer hardware in the form of tablets, laptops, or desktops. However, most schools in marginalised communities have challenges getting funding to ensure a 1:1 computer-learner ratio and they rely on government funding [24]. Where government funding is insufficient, these schools need to search for alternative sources to stock up their computers. Another strategy for ensuring adequate computers in schools is the use of the Bring Your Own device strategy where learners provide their own gadgets [25]. The challenge with this strategy comes with regards to learners who cannot afford these gadgets. In addition, the issue of cost even goes beyond acquiring computers in schools. The running costs of technology include purchasing software, maintenance, replacement parts, insurance, and internet access [26]. Marginalised schools, particularly those in rural areas of most developing countries are affected by a lack of electricity to power the computers, safe storage space and general infrastructure [26]. These can prove costly and result in non-implementation of technology in under-funded institutions.

Another challenge of note in the implementation of technology in the education sector is the alienation of socio-economic groups. [24] reported that in South Africa, although there is a significant increase in participation in technology-driven learning areas, a clear divide is shown in the proportion of white students and that of African descent. This is a direct influence of the colonial era which saw white students having a historical advantage over those of African descent and this has translated into economic status and participation in technology-based subjects or learning areas [27].

Although the provision of Information and Communication Technology (ICT) infrastructure is detailed in the policies of many countries, debatably, many states have not prioritised it. Many African governments have comprehended it as a secondary need, especially with regards to marginalised areas where ICT has not been introduced in the first place. In Kenya, the Poverty Reduction Strategy has acknowledged the role of technology in poverty alleviation. However, 'bureaucracy, a lack of professionalism in the business sector, corruption, illiteracy, and poverty' have hampered the roll-out of ICT programmes in different sectors including schools [28].

After the availability of computers, learners should have enough contact time with the computers. Even though many schools are working towards a 1:1 computer-learner ratio, effective learning only occurs if they are given adequate time to access the computers. [29] revealed that the minimum recommended period should be an hour of contact time. In addition to adequate contact time, consistency and regularity should be maintained.

Training is also a huge constraint in the implementation of technology in schools. Teachers must be technologically trained so that they will be able to impart skills to learners. In some instances, teachers will be constrained from fully harnessing the potential of technology in the classroom. The biggest constraint to incorporating technology into teaching and learning is insufficient teacher training [30]. Teacher training cultivates confidence in using teaching and learning technology, surfing the internet during research, or using various software. Additionally, teacher training is not a once-off event, thus frequent training is a necessity since innovation and technology are ever-changing [27]. As a result, schools and responsible authorities should constantly offer capacity-building exercises to boost their teachers' confidence and skills. Even teachers who are appointed who have already existing computer skills need capacity building because of the constant technological changes.

After the provision of computers and training, support barriers follow. Technology being ongoing, various kinds of support are needed both in the initial phases and many times after. In this respect, support comes through technical support and administrative/peer support [27]. With technical support, content creators of educational software/programmes are expected to give continuous support to users. Support can also be rendered through peer groups as the level of expertise increases. On the other hand, a lack of support from content creators will result in the failure of technology to be implemented in schools.

Internal barriers or second-order barriers to implementing technology in the classroom involve the decisions that teachers make to use technology in their teaching [31]. These barriers include attitude, beliefs, and knowledge [27]. The attitude of the teacher on pedagogy and the incoming technology determines the extent of use and incorporation into the curriculum. While well-equipped schools focus on how technology can be incorporated into teaching and learning, ill-resourced and marginalised schools focused on whether technology should be included [30]. These two approaches result in differences in the level of incorporation into teaching and learning. The latter represents schools where traditional teaching methods are upheld at the expense of technology. [26] reported that a lack of know-how on the use of technology is one of the major factors affecting the implementation of technology in rural schools in South Africa.

Second-order challenges are also presented through confidence in skills and knowledge. Most of the learners nowadays belong to a class of 'digital natives' who grew up with technology and are familiar with it [32]. Most teachers are referred to as 'digital immigrants' who were born without digital exposure and were introduced to it at a later stage [32]. Due to this dynamic, most teachers may feel that they are not technologically competent as opposed to the learners whom they are supposed to teach, and this may reduce their confidence levels. This lack of confidence will strip them of control in the classroom, hence they will stick to the traditional methods of teaching, disregarding the role of technology in teaching and learning [27].

1.3 Technologies to Monitor the Girl Child at School

a. Bluetooth-powered messaging service

According to [33] the Centre for Social Concern and Development (CESOCODE), a non-governmental organisation based in Malawi was monitoring the welfare of girls and assisting with necessary interventions during the pandemic, utilising technology as the main intervention tool. The organisation used bluetooth-powered mobile-to-mobile messaging service to monitor the girls during the time they were out of school due to lockdowns. The girls can report any cases of abuse through this service.

b. Social media

CESOCODE has also made use of social media platforms such as Facebook to broadcast podcasts focusing on educating and conscientising girls on issues about domestic violence and health. The podcast is also delivered in Sign Language to cater for girls with hearing disabilities. The girls can report any abuse or challenges they are facing using social media such as WhatsApp and Facebook [33].

c. Loudspeakers

CESOCODE also educates girls in rural areas through messages broadcasted by loudspeakers. A vehicle moves around communities, with someone educating the masses through loudspeakers. They also make use of posters, brochures, and flyers to disseminate messages of hope to the girls [33].

d. Use of helpline services

Telephone helplines originated around the 1950s and their popularity was exacerbated by the increase in the use of the telephone as a mode of communication [34]. The first helplines to be established were by Samaritans and Childline. Helplines offer education and support for medical and psychological conditions to reduce a wide array of challenges in communities. Currently, different kinds of helplines are available, including Human Immunodeficiency Virus (HIV), child protection, girl-child protection, suicide prevention, domestic abuse, sexual abuse, addiction, depression, sexual identity, and bullying [35, 36]. Helplines offer different scopes of assistance to the callers. Some offer counsel and information while others are contact points for reporting social ills. In addition, helplines are available free of charge and currently, calls can be made from any type of phone, whether landline or cell phone. Helplines have also diversified their modes of contact into emails, text messages, video-calling and chat services [37]. Most helplines operate 24 h daily to ensure accessibility. Helplines can be used to ensure the well-being of girls such that they will be in the right state of mind to attend school and be as competitive as boys of their age.

2 Description of the Case

The Eastern Cape Province is one of the most negatively affected provinces in terms of learner dropout rates. The province is dominated by townships to rural areas. There is no doubt that those who live below the poverty datum line are always hit harder by any

challenge. The focus is on girl children as they are the ones who are gravely affected with dropping out. While this disruption to education and the expected reduction in global growth have far-reaching effects for all, their impact will be particularly detrimental to the most disadvantaged students and their families, especially in poorer countries [38]. Education across the globe faces great challenges, and developing countries such as South Africa are gravely affected, by socio-economic factors, infrastructure challenges, and a lack of resources. Approximately, 300 000 children have potentially dropped out of primary schools across South Africa in 6 months, including the national lockdown. Then, 126,553 children in KwaZulu-Natal have missed school, 114, 558 in Western Cape, 8153 in Eastern Cape, 55,571 in Gauteng and 800 in Limpopo [39].

Eastern Cape province, the poorest province in South Africa, has recorded approximately 130,000 learner dropouts during the 2020 academic year at all levels. As Covid-19 forced 743 million girls out of school in 185 countries, the rising dropout rates disproportionately affected adolescent girls, as disease outbreaks affect girls and boys differently. This exacerbates gender gaps in education and leads to increased risk of sexual exploitation, early and unintended pregnancy, and child, early and forced marriages [40].

In South Africa, most school dropouts take place in Grades 10 and 11. Approximately, 60% of young South Africans dropout of school without finishing Grade 12 and obtaining their matric certificate, which is the school leaving qualification. There is a huge inequality gap in South Africa, and this can also be identified in the levels of dropout rates [12]. As a result of the country's racist, colonial and apartheid social and economic policies, the levels of dropout differ significantly by race [41]. According to a 2011 General Household Survey, the attainment of matric was so imbalanced, with only 44% of Black and Coloured youth aged 23–24 attaining matric compared to 83% of Indian youth and 88% of White youth [42].

2.1 The OR Tambo District

Learners who are mostly absent at school are at risk of problems associated with drug and alcohol abuse, and personality changes turn to crime to support expensive chemical dependence habits. Worth noting girls also bear a greater negative cost with rises in sexual abuse, teenage pregnancy, and early marriage. The most concerning of them all is the fact that distance learning solutions, whether low- or high-tech, often indirectly discriminate against girls due to power dynamics within families. Besides, when schools re-open, poorer children and girls are less likely to return, with increased dropout rates caused partly by fear, stigmatisation and, in some cases, deliberate exclusion.

This is a result of the difficulty in attempting to conceptualise how each of the complex risk factors, such as family, school, individuals, and community interact during the process of disengagement and dropout [43]. A single risk factor cannot be used when an accurate prediction of dropout cause, risk or impact is made. The OR Tambo district as shown in Fig. 1 is the second district with the highest number of high school dropouts.

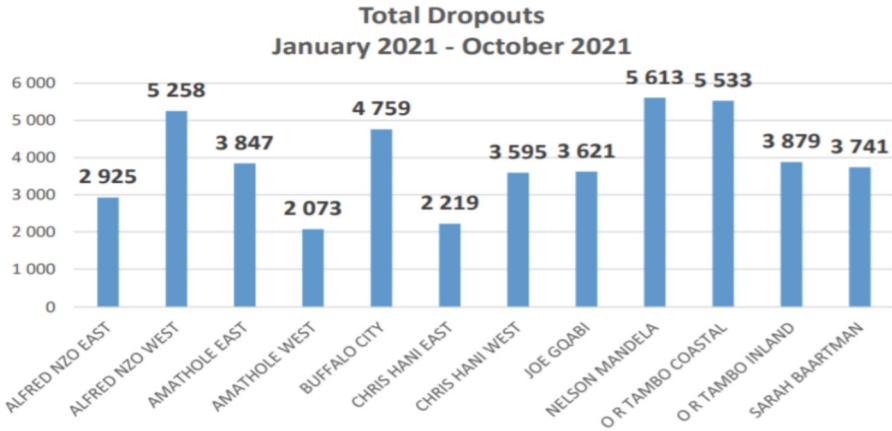


Fig. 1. Eastern Cape District’s Dropout Statistics

3 Girl Child School Dropout Model

The big data analytics tool was developed through the data collected in the selected schools for this study. The tool will assist in identifying potential trends in the behaviours of the girl child, thus assisting in the formulation of situation-specific intervention strategies. The analysis of data and the implementation of machine learning algorithms to predict dropouts entails scrutinising patterns, trends, and risk factors among students who are at risk of dropping out of school [44]. The Web-Based Big Data analytics tool is designed for girls in high schools across Eastern Cape province to connect the capabilities of Big Data to predict the likelihood of dropout. The tool can be utilised by various stakeholders in the Department of Basic Education to predict the dropping out of a girl child. Users of the tool in various schools include principals, teachers, and the girlLearners. The tool’s predictive capabilities can be utilised as an intervention tool to pinpoint students who are at risk of discontinuing their education. This identification process can be based on a multitude of factors, such as attendance, academic performance, and socio-economic background [45]. A range of analytical techniques such as trend analysis have been applied in dropout prediction, including the examination of historical data by analysing past student records, academic performance, attendance patterns, and demographic information [44]. Once learners who are at risk are identified, schools can thereafter implement proactive interventions. These interventions can take the form of personalised counselling, mentorship programmes, or additional academic support to address specifically the unique challenges faced by each student [45]. The integration of a predictive tool for education within a smart city framework can contribute to the overall development and efficiency of the city. By analyzing student enrollment patterns, demographic changes, and historical data, it is possible to predict future education resource needs. These insights can be utilized by smart cities to optimize the allocation of resources, including classrooms, teachers, and educational materials, based on projected demand. This ensures that educational facilities are efficiently distributed (Fig. 2).

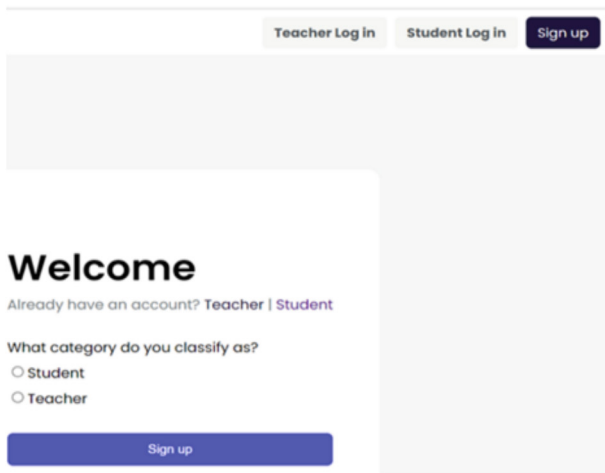
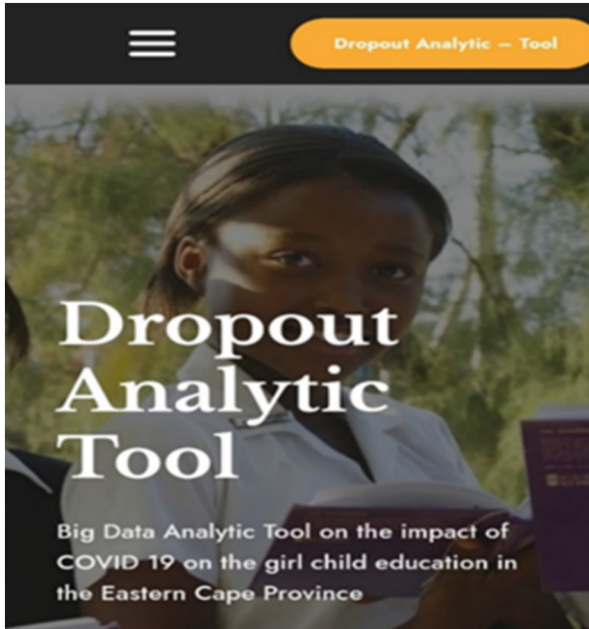


Fig. 2. Drop out Predictive Tool Home Page

4 Methodology

Primary and secondary data were collected. Data were collected using qualitative and quantitative methods. Qualitative data were obtained from interviews, focus groups and participatory design. The collected data were recorded and transcribed to work out related topics. The study collected data on female learners from schools in OR Tambo district in 2021. Over 600 schooling girls participated. The questionnaires were distributed systematically and randomly to students in 10 schools in OR Tambo district. For qualitative data, all collected data from focus groups, workshops and interviews were transcribed and key themes were formulated. The common dropout factors from the various participants were grouped and then a ranking was provided. Various stakeholders from the 10 participating schools were involved. These included parents, teacher, learners, principals, and Eastern Cape District Officials. Only data collected from the School Governing Body, principals and teachers in OR Tambo District were considered for this paper.

- The specific data collection tools and methods included:
- Workshops
- Participatory design
- Focus groups
- Interviews
- Experiential and prototype design
- Surveys – online and face-to-face

4.1 Qualitative Results

For the qualitative approach, all stakeholders were engaged through focus groups, workshops, and in-depth interviews to elicit the causes of dropout among learners. The dropout factors were ranked and considered for developing a big data analytic tool that assists in predicting that a learner is likely to be a dropout.

Data were collected from 10 selected schools. Five schools were inland and five from the coastal region of OR Tambo district. The principals were engaged through one-on-one interviews. The rest of the stakeholders, including teachers and SGBs were engaged through focus groups and workshops.

a) Principals

Learner dropout has always been a prevalent challenge in schools, especially in rural areas, where access to resources and schools itself is difficult. The school principals interviewed in Eastern Cape identified teenage pregnancy as the main cause of girl-child abandonment,.

Below are the views of some of the principals:

“The school is in one of the poorest areas, you find that some of the kids have parents who are not working, in fact, in the whole homestead, no one is working. You find that in some of the children’s homesteads, the parents are not around, they have migrated to other areas to try find employment, leaving the children alone to be independent”.

“At one time, a learner experienced contractions at school and she had to be rushed to hospital with our cars so after that the school requested that parents be

present in class with their children when they are pregnant so that they are the ones who take care of them when giving birth. This is why some parents decide to keep their pregnant children at home because they cannot be on guard 24/7”.

“When boys go to initiation, which is seasonal, you find that some of them go but do not come back, which also ultimately leads to dropping out, and this may be caused by various reasons, although this study is for girls, boys dropping out sometimes leads to girls also dropping out since some of them follow their boyfriends, get married and have children.”

Principals pointed out that another important factor is social harm in the communities in which they serve. Students are without their biological parents due to death or distance, which means that they either stay with grandparents or other family members or they stay alone to become victims of crime, rape and are caught in blesser-blessee relationships for their survival. They experiment with drugs, are involved in unprotected sex, fall pregnant, and have no support after giving birth, leading to abandonment.

Ratings of factors provided by the principals:

- Top 3 Ranked: Pregnancy, Drug and Substance Abuse and Child-Headed Homes.
- 3 Medium Ranked: Socio-Economic Factors, Peer Pressure, and Rape.

b) School Governing Body

In this study, the parents revealed that the dropout rate of girls is higher than the dropout rate of boys and the foremost cause for girls to dropout is pregnancy.

Below are the views of some of the SGB:

“Mothers have a big contribution in these girls’ pregnancies because if they get plastics of groceries from their daughters’ sugar dads, they allow them to do whatever they like”.

“Akunkwenkwe akuntombazana, bayanxila qha, and we cannot say anything because siyaboyika xa bekulameko”

Translated as *“Whether it’s a boy or a girl, they’re just drinking, and we cannot say anything because we fear them when they are in that intoxicated state”.*

Ratings of factors provided by the School Governing Body:

- 3 Top Ranked: Pregnancy, Drug and Substance Abuse, and Repeating Grades.
- 3 Medium Ranked: Socio-Economic Factors, Child-Headed Homes, and Learner Rotation in Schools Due to Covid-19 (Fig. 3).



Fig. 3. School Governing Body's Focus Group Discussions

c) Teachers

The teachers were asked to give their opinions on the reasons for the girl children dropping out of school. They pointed to several factors that contribute to girls dropping out of school. The teachers mentioned that the predominant factor is teenage pregnancy, followed by drug and alcohol abuse. They believed that female children have been subjected to drug abuse, particularly dagga as these are readily available. Another noticeable trend was teenage pregnancy caused by teachers themselves.

Below are the views of some of the teachers:

“There are a lot of cases, as men we play a big role towards schoolgirls, for example, a male teacher can sleep with a learner; and when the girl becomes pregnant, the teacher can decide to deny the child, then, the girl decides to do abortion at school, and it will be known that that girl was impregnated by the teacher, thus the girl ends up not going to school”.

“Poverty is the main cause of dropout because a learner cannot afford a new uniform, as most of the time, other learners are given school uniform by former learners. As time goes on, the uniform is torn and a student cannot further her studies because she does not have a uniform. Financial instability, not having money to pay rent as most learners are renting when they are doing Grade 12, most students in their home survive on grants; most of the parents cannot afford to pay rent for their learners to be close to school so other learners decide to dropout because of lack of financial support.”

“They are also a lot of rape cases around our community in which girls are raped and threatened so most of the girls end up leaving school because of traumatic experiences.”

“Pregnancy is one of the factors that cause school dropout, as when girls are pregnant, their parents do not encourage them to go back to school after giving birth.”

“The only monitoring system is the register.”

Ratings of factors provided by the teachers:

- 3 Top Ranked: Drug and Substance Abuse, Pregnancy, and Socio-Economic Factors.
- 3 Medium Ranked: Rape, Parents Deceased Due to Covid-19, and a Lack of Transport.

5 Analysis of Drop-Out Factors

Intrinsic Factors: related to learners’ individual dispositions, behaviours or personal choices that lead to early school leaving. Student Individual: Traits students possess regardless of demographics such as substance abuse, law violations, and pregnancies. *Push out:* school-related factors that are experienced by the learners as exclusion from school. These included both unintentional and conscious efforts by educators or school management to limit the number of learners who write Grade 12 community examinations, which lead to a student dropping out of school.

Extrinsic: barriers that exist at home or in the community that have resulted in a learner dropping out of school

The web-based big data analytics tool was developed for girls in high schools in Eastern Cape province to connect the power of big data to predict the likelihood of dropping out of school. The tool can be used by the various stakeholders in the Department of Basic Education to predict a girl’s dropout rate. The users of the tool in the different schools include principals, teachers, and students. The big data analysis tool was developed using the data collected in the schools selected for this study. The tool helps identify potential behavioural trends in girls, thereby helping in formulating situation-specific intervention strategies.

The tool is useful for the education sector, but the information and topics covered are unlimited to the sector. Information on the interconnected socio-economic, environmental, political, and cultural challenges is discussed on the tool’s homepage. The learner is asked a series of questions to learn more about the challenges, life circumstances, financial situations and various social issues that apply to a learner (Fig. 4).

The image shows a mobile application interface for a predictive tool. It consists of several sections:

- Survey Questions:** Four questions are listed, each with a blue header box and radio button options for 'Yes' and 'No'.
 - Question 1: "Is there anyone who is bullying you?" with radio buttons for "Yes" and "No".
 - Question 2: "How would you rate the affordability of your Parents". The text below the question reads: "1 if they cannot afford - 10 they can afford anything I need (Stationery, school uniform, school fees)". Below this is a slider control with a blue knob and the number "10" at the right end.
 - Question 3: "Do you have a boyfriend?" with radio buttons for "Yes" and "No".
 - Question 4: "Are you sexually active?" with radio buttons for "Yes" and "No".
- Result Notification:** A red banner at the bottom of the survey section contains the text: "Your result indicates that you are at a [HIGH] risk of dropping out".
- Action Button:** A blue button labeled "Check my status" is positioned below the red banner.
- Status Review Section:** A large white box with a grey border contains the heading "My Status Review" and a sub-heading "Age". Below "Age", the text reads: "Your age is above the grade you are in currently."

Fig. 4. Predictive Tool's User Interface

Because of personal questions learners are asked, the tool makes provision for privacy controls to protect the learners' rights to privacy. Teachers and principals are unable to see the learners' responses to all the questions asked on the system.

All the questions asked on the tool are rated as per the results from the data collected within the schools and ratings from the stakeholders as well as the literature consulted. After answering all questions asked, the learner can click on the [check status] button, and their results will be displayed (Fig. 5).

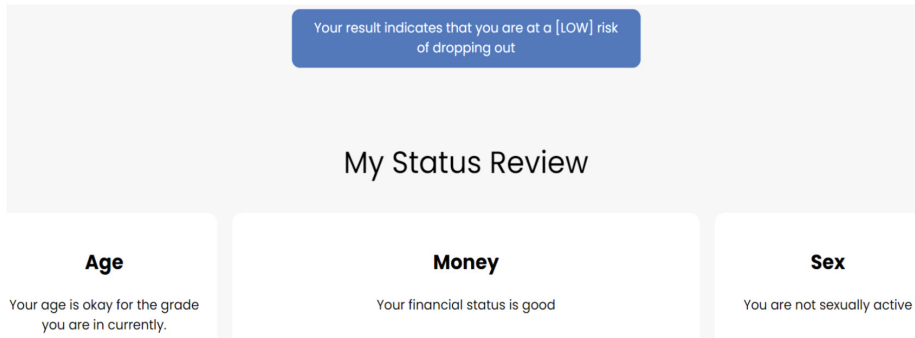


Fig. 5. Dropout Prediction Status Review

The tool allows teachers/principals to utilise the information for report-based purposes. After a teacher signs up on the tool, they have only view permissions of the students' final prediction results. The teacher can only view the level of dropout risk for a learner. The learners' personal details and responses are only visible to them.

6 Conclusion

There are large quantities of data in the education sector, and this study demonstrated this through only one district in the province which contributed towards the dropout prediction tool. The ultimate goal is to connect the capabilities of Big Data to accelerate and enhance the impact of the sector and provide solutions to current and future problems encountered as a result of learner dropout. Through the data analytics capabilities of the tool, there is a lot that has been discovered, mostly through the help of teachers, parents and principals who contributed greatly by sharing their real-life experiences of the education system and rating dropout factors. This tool was developed based on the ratings from all the participants and is key to expanding further studies across other districts and eventually, the entire Eastern Cape Province. The study shows that understanding of the existing digital inequalities is critical towards innovative solutions. Data analytics based on societal factors is relevant in ensuring smart sustainable solutions for South Africa.

References

1. Sarker, Md., Islam, N., Min, W., Hossin, M.A.: Economic effect of school dropout in Bangladesh. *Int. J. Inform. Educ. Technol.* **9**(2), 136–142 (2019)
2. Mennen, O.K., Weybright, W.C., Aldridge, N.S.: Effects of child abuse on school dropout among students in USA. *J. Educ.* **5**(2) (2022). <https://doi.org/10.53819/81018102t5065>
3. Fleisch, B., Shindler, J., Perry, H.: Who is out of school? Evidence from the Community Survey 2007, South Africa (2010)
4. Meinck, F., Cluver, L.D., Boyes, M.E., Ndhlovu, L.D.: Risk and protective factors for physical and emotional abuse victimisation amongst vulnerable children in South Africa. *Child Abuse Rev.* **24**, 182–197 (2015)
5. Tulasi, B.: Significance of big data and analytics in higher education. *Int. J. Comput. Appl.* **68**(14), 21–23 (2013)
6. Fatima, S.A., Fatima, S.T., Alwi, S.K.K.: Major causes of female dropouts at different educational levels in Karachi. *Global Educ. Stud. Rev.* **VI**(I), 293–305 (2021). [https://doi.org/10.31703/gesr.2021\(vi-i\).30](https://doi.org/10.31703/gesr.2021(vi-i).30)
7. Job, M.: An efficient way of applying big data analytics in higher education sector for performance evaluation. *Int. J. Comput. Appl.* **180**(23), 25 (2018)
8. Yukselturk, E., Ozekes, S., Türel, Y.: Predicting dropout student: an application of data mining methods in an online education program. *Eur. J. Open Distan. e-Learn.* **17**(1), 119 (2014)
9. Magana, A.J., Silva Coutinho, G.: Modeling and simulation practices for a computational thinking-enabled engineering workforce. *Comput. Appl. Eng. Educ.* **25**(1), 62–78 (2017)
10. Khan, M., Ibrahim, R., Ghani, I.: Cross domain recommender systems: cross domain recommender systems. *ACM Comput. Surv.* **50**(3), 1–27 (2017)
11. Verma, A., Kumar, D.: Evaluating and enhancing efficiency of recommendation system using big data analytics. *Int. Res. J. Eng. Technol.* **4**(3), 1518–1527 (2017)
12. Motala, S., Dieltiens, V., Sayed, Y.: Physical access to schooling in South Africa: mapping dropout, repetition and age-grade progression in two districts. *Comp. Educ.* **45**, 251–263 (2009)
13. De Ridder, K.A.A., et al.: High school dropout and long-term sickness and disability in young adulthood: a prospective propensity score stratified cohort study (the Young-HUNT study). *BMC Public Health* **13**(1) (2013). <https://doi.org/10.1186/1471-2458-13-941>
14. Nabugoomu, J.: School dropout in rural Uganda: stakeholder perceptions on contributing factors and solutions. *Educ. J.* **8**(5), 185 (2019). <https://doi.org/10.11648/j.edu.20190805.13>
15. Nabugoomu, J., Seruwagi, G. K., Hanning, R.: Addressing poverty, malnutrition and poor health for adolescent mothers in rural eastern Uganda: recommendations of local level stakeholders. *Afric. J. Health Sci.* **33**(4) (2020)
16. Nakajima, M., Kijima, Y., Otsuka, K.: Is the learning crisis responsible for school dropout? a longitudinal study of Andhra Pradesh, India. *Int. J. Educ. Dev.* **62**, 245–253 (2018). <https://doi.org/10.1016/j.ijedudev.2018.05.006>
17. Ramsdal, G.H., Wynn, R.: How young people who had dropped out of high school experienced their re-enrolment processes. *Int. J. Educ. Res.* **106** (2021). <https://doi.org/10.1016/j.ijer.2021.101732>
18. UNESCO. WHAT'S NEXT? lessons on education recovery: findings from a survey of ministries of education amid the COVID-19 Pandemic (2021). <https://en.unesco.org/gem-report/taxonomy/term/238>
19. Onyema, E.M., et al.: Impact of coronavirus pandemic on education. *J. Educ. Pract.* (2020). <https://doi.org/10.7176/jep/11-13-12>

20. Laura, M., Evans, D.K.: Learning Loss and Student Dropouts during the COVID-19 Pandemic: A Review of the Evidence Two Years after Schools Shut Down. CGD Working Paper 609. Center for Global Development, Washington, D (2022). <https://www.cgdev.org/publication/learning-loss-and-student-dropouts-during-covid-19-pandemic-review-evidence-two-years>
21. Manasia, L., Ianos, M.G., Chicioareanu, T.D.: Pre-service teacher preparedness for fostering education for sustainable development: an empirical analysis of central dimensions of teaching readiness. *Sustainability (Switzerland)* **12**(1), 4–6 (2020)
22. Mhlanga, D.: Industry 4.0: the challenges associated with the digital transformation of education in South Africa. In: Aydin, O. (ed), *the impacts of digital transformation*, efcademy, Instabul, pp. 12–25 (2020)
23. Fomunyam, K.G.: Education and the fourth industrial revolution: challenges and possibilities for engineering. *Int. J. Mech. Eng. Technol.* **10**(08), 271–284 (2019)
24. Kayembe, C., Nel, D.: Challenges and opportunities for education in the Fourth Industrial Revolution. *Afric. J. Public Affairs* **11**(3), 79–94 (2019)
25. Afreen, R.: Bring Your Own Device (BYOD) in higher education: opportunities and challenges. *Int. J. Emerg. Trends Technol. Comput. Sci.* **3**, 233–236 (2014)
26. Dzansi, D.Y., Amedzo, K.: Integrating ICT into rural South African schools: possible solutions for challenges. *Int. J. Educ. Sci.* **6**(2), 341–348 (2014)
27. Johnson, A.M., Jacovina, M.E., Russell, D.E., Soto, C.M.: Challenges and solutions when using technologies in the classroom. In: Crossley, S.A., McNamara, D.S. (eds.) *Adaptive educational technologies for literacy instruction*, pp. 13–29. Taylor & Francis, New York (2016)
28. Hallberg, D., Kulecho, M., Kulecho, A., Okoth, L.: Case studies of Kenyan digital villages with a focus on women and girls. *J. Lang. Technol. Entrepren. Africa* **3**(1), 255–273 (2011)
29. Warschauer, M., Zheng, B., Niiya, M., Cotten, S., Farkas, G.: Balancing the one-to-one equation: Equity and access in three laptop programs. *Equity Excell. Educ.* **47**(1), 46–62 (2014)
30. Ertmer, P.A., Ottenbreit-Leftwich, A., Sadik, O., Sendurur, E., Sendurur, P.: Teacher beliefs and technology integration practices: a critical relationship. *Comput. Educ.* **59**, 423–435 (2012)
31. Ertmer, P.A.: Addressing first-and second-order barriers to change: Strategies for technology integration. *Educ. Tech. Res. Dev.* **47**(4), 47–61 (1999)
32. Gallardo-Echenique, E., Marques-Molas, L., Bullen, M., Strijbos, J.: Let's talk about digital learners in the digital era. *Int. Rev. Res. Open Distrib. Learn.* (2015). <http://www.irrodl.org/index.php/irrodl/article/view/2196/3337>. Accessed 23 Oct 2021
33. Carroll, R.: The new normal- how technology is helping CESOCODE prevent child marriages during COVID (2020). <https://www.girlsnotbrides.org/articles/the-new-normal-how-technology-is-helping-cesocode-fight-child-marriage-during-covid-19/>. Accessed 7 Oct 2021
34. Morgan, K., Chakkalackal, L., Cyhlarova, E.: Lifelines: Evaluation of mental health helplines. Mental health helpline partnership (2012). https://www.mentalhealth.org.uk/sites/default/files/life_lines.pdf. Accessed 14 Oct 2021
35. Arullapan, N., et al.: Quality of counselling and support provided by the South African National AIDS Helpline: content analysis of mystery client interviews. *S. Afr. Med. J.* **108**(7), 596–602 (2018)
36. UNFPA–UNICEF. Adapting to COVID 19 (2020). https://www.unfpa.org/sites/default/files/resource-pdf/Responding_to_COVID-19_Pivoting_the_GPECM_to_the_pandemic.pdf. Accessed 14 Oct 2021
37. Dinh, T., Farrugia, L., O'Neill, B., Vandoninck S., Velicu, A.: Insafe helplines: operations, effectiveness and emerging issues for internet safety helplines (2016). <https://www.betterinternetforkids.eu/documents/167024/507884/Helpline+Fund+Report+-+Final/4cf8f03a-3a48-4365-af76-f03e01cb505d>. Accessed 14 Oct 2021

38. De Witte, K., Cabus, S., Thyssen, G., Groot, W.: A critical review of the literature on school dropout. *Educ. Res. Rev.* **10**, 13–28 (2013)
39. Mhlanga, D., Moloji, T.: COVID-19 and the digital transformation of education: what are we learning on 4IR in South Africa? *Educ.Sci.* **10**, 180 (2020). <https://doi.org/10.3390/educsci10070180>
40. Hutchinson, G.: How will Plan International support girls, young women and vulnerable groups? s.l.: s.n (2020)
41. Moses, E., van der Berg, S., Rich, E.: A society divided: how unequal education quality limits social mobility in South Africa. RESEP: University of Stellenbosch.: Synthesis report for the Programme to Support Pro-poor Policy Development (PSPPD) (2017)
42. Hartnack, A.: Background document and review of key South African and international literature on school dropout. Sustainable Livelihoods Foundation, Cape Town (2017)
43. Hallgarten, J.: Four lessons from evaluations of the education response to Ebola, s.l.: Education Development Trust (2020)
44. Kawchale, N., Satao, R.: Predictive analytics in education context. *Int. J. Mod. Trends Sci. Technol.* **02**(11), 77–82 (2016)
45. Biasi, V., De Vincenzo, C., Fagioli, S., Mosca, M., Patrizi, N.: Evaluation of predictive factors in the drop-out phenomenon: interaction of latent personal factors and social-environmental context. *J. Educ. Soc. Res.* **9**(4), 92–103 (2019). <https://doi.org/10.2478/jesr-2019-0059>