AN EARLY WARNING SYSTEM BASED ON PROBABILISTIC DISTANCE CLUSTERING ALGORITHM FOR STUDENT AT RISK DETECTION

Fezile Matsebula and Randle Oluwarotimi

“The term 'early warning' is used in many fields to describe the provision of information on an emerging dangerous hazard that enables advance action to reduce the associated risks”

Purpose of an EWS

The goal of the project is to develop an early warning system that effectively identifies students “at risk” of dropping out.

To use data to make an informed decisions regarding the identification of students with service monitor progress

The ability to identify and describe, those students who are at greatest risk of academic failure so that intervention can occur early.

Allowing institutions to confidently develop and implement appropriate intervention programming.

The primary function of an early warning system is to alert academics, parents, and students when a student falls off track

PROJECT PHASES

Phase 1:

- Conduct literature Review
- Establish teams
- Develop a work plan

- **Phase 2:**
- Collect system requirements
- Develop data system
- Conduct preliminary screening

Phase 3:

Develop an intervention and monitoring system

PHASE 1

Which students are most likely to dropout?

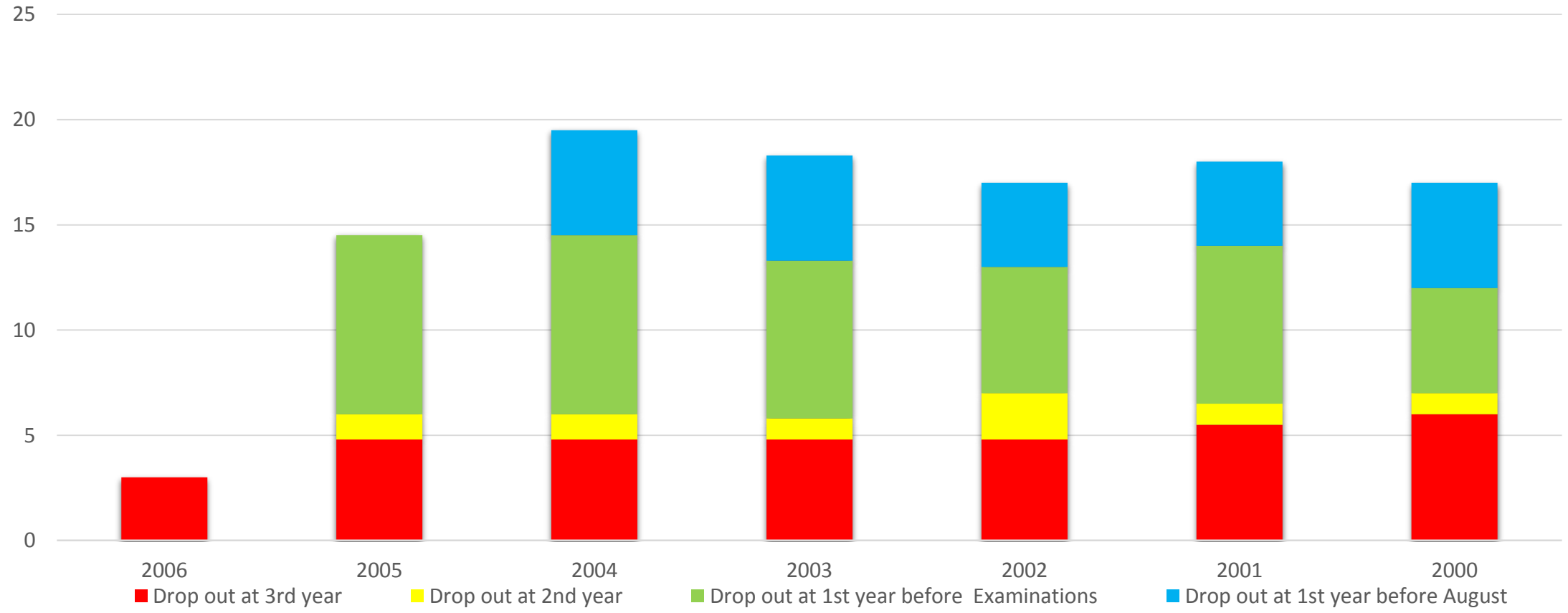
How early can we identify those students at risk?

Can we identify students at risk by analysing data from LMS or do we need specialized assessments?

THE SOUTH AFRICAN HIGHER EDUCATION LANDSCAPE

- Department of Education reported that of the 120 000 students who enrolled in higher education in 2000, 36000 (30%) dropped out in their first year.
- A further 24000 (20%) dropped out during their second and third year.
- Of the remaining 60 000, 22% graduated within the specified three-year duration for a generic Batchelors degree
- This dropout cost the National Treasury R4.5 billion in grants and subsidies to higher education institutions without a return on investment

How early can we identify those students at risk?



ESTABLISHING A TEAM

- Teaching and learning
- Financial Aid
- Academic support and advising
- Student Accounts
- Admissions
- University Senate

Which students are most likely to dropout?

- Student absence
- Poor career guidance
- Low academic achievement
- Transition
- Low socio-economic status
- Behavioural problems

PHASE 2: DEVELOP DATA SYSTEM

What makes up a good early warning system (characteristics)?

Where will the data be harvested?

What resources are needed to capture the signal: Technological and organizational?

Characteristics of an early warning system

- **Accessibility of data:** Data that will be needed to identify students at risk must be readily available and accessible.
- **High accuracy**:
  - higher percentage of students with the “signals”- drop out.
  - lower percentage of students without the “signals” graduate.
- **Empirically developed:** Through analysis of longitudinal data for prior cohorts of students, signals will identified.

Data house

- Pre-enrolment data
- Assessment test
- First year academic achievement

PDC – The Ranking Algorithm

```
Insert Student Name and 3 Values: tim 34 54 65
Insert Student Name and 3 Values: wen 45 65 78
Insert Student Name and 3 Values: jim 45 98 90
Insert Student Name and 3 Values: ral 98 87 98
Insert Student Name and 3 Values: zil 98 34 54
Insert Student Name and 3 Values: pal 87 65 87
Insert Student Name and 3 Values: edd 98 43 54
Insert Student Name and 3 Values: que 23 43 23
Insert Student Name and 3 Values: tum 32 78 78
Insert Student Name and 3 Values: set 98 56 87
Insert Student Name and 3 Values: ted 87 76 87
Insert Student Name and 3 Values: ile 89 64 34
Insert Student Name and 3 Values: sed 76 54 65
Insert Student Name and 3 Values: den 67 76 85
Insert Student Name and 3 Values: fed 87 75 69
Insert Student Name and 3 Values: olu 89 54 63
Insert Student Name and 3 Values: ade 78 76 57
Insert Student Name and 3 Values: red 87 67 43
Insert Student Name and 3 Values: ted 64 54 34
Insert Student Name and 3 Values: que 76 54 56
```

```
den 0.999992
ade 0.999985
fed 0.999984
ted 0.999983
ven 0.999982
tun 0.999981
pal 0.999976
ral 0.999976
jin 0.999975
red 0.999973
sed 0.999973
ted 0.999972
que 0.999972
tin 0.999971
olu 0.999969
ile 0.999968
set 0.999965
edd 0.999955
que 0.999955
zil 0.999948
```

What resources are needed to capture the signal: Technological and institutional?

- Innovative approaches to improve student at risk
- Potential predictive models
- Teams to collate all data from various domains
- Learning Analytics Centre

PHASE 3: INTERVENTION AND MONITORING SYSTEM

- **Before the beginning of studies**
  - Identify students who are at risk
  - Students must be assigned to an advisor
  - Peer tutoring

- **After they have started**
  - Teaching approach to critical courses
  - What happens in the classroom – lecturer/student
  - Quality students vs Dropout rates – high quality research-based instruction and behavioural support
  - Progress monitoring

What happens to students who fail their courses?

- Do we let them dropout or redirect them to another course?
- Do we examine the program admission criteria?
- Do we revisit or restructure the institutional process?