Kelle Clark’s Proposed Outline (Slides) for Project (Slides) “Data Mining the Wingate Mathematics Placement Test for Knowledge Discovery”

1. What is the Information Need?

<BRIEFLY>

Issue: Math placement for incoming students into first math course…

Known correlation between success course and previous entrance SAT/ACT admission levels  
have new entrance standards…need more criteria/guidance.

Current measuring tool – Math Placement Test ---show questions.

Appendix contains placement test

What is known so far about these tools and the placement – what do we

need to address, info needed?

Reference of Wingate Math department internal paper summarizing results of data at hand before this study.

1. What considerations were involved when designing the project?

<BRIEFLY> describe

LONG process of applying for RRB approval….

demonstrate what these forms looks like

Online class required through Citi…

show sample of topics and why this is of concern

flash certificate earned

1. How was data collected and what data was collected?

Discuss variables and types:  
 DATASET “Math Placement Results – 2017 – 2019”

Semester\_taken {FALL, SPRING, SUMMER}  
 ID Used to match with other dataset

Math Recommendation {MATH116, MATH117, MATH120,MATH209,MATH112,MATH115}

Adjusted\_score. {0,…,2040}

Score {0,40}

DATASET “Alldata”

Academic\_year {? Represent dates in Weka?}

Academic\_term {FALL, SPRING, SUMMER}

Event\_ID {MATH116, MATH117, MATH120,MATH209,MATH112,MATH115}

Start\_date {? Represent dates in Weka?}

Person\_code\_id. Used to match with other dataset

Status\_date {? Represent dates in Weka?}

Last\_year {? Represent dates in Weka?}

Last\_term {FALL, SPRING, SUMMER}

Mid\_grade {A+, A, A-, B+, B, B-, C+, C, C-, D+, D,D-, F, W}

Final\_grade {A+, A, A-, B+, B, B-, C+, C, C-, D+, D,D-, F, W} 🡨 class

\*student ids are randomized labels and do not identify student

I would like to follow something similar to the techniques used in the following MS Thesis: A screenshot of a social media post

Description automatically generated

1. Preprocessing –

Understanding the distribution of each variable

Combining different datasets

Normalization, discretization, clustering

1. What types of associations are suggested by the data?
2. Build classifiers using a n-fold approach on data
3. Assess/Evaluate models
4. What did we learn?