Predicting Academic Performance: A Systematic Literature Review

Arto Hellas
University of Helsinki
Helsinki, Finland
arto.hellas@cs.helsinki.fi

Petri Ihantola
University of Helsinki
Helsinki, Finland
petri.ihantola@helsinki.fi

Andrew Petersen
University of Toronto Mississauga
Mississauga, Canada
petersen@cs.toronto.edu

Vangel V. Ajanovski
Saints Cyril and Methodius University
Skopje, Macedonia
vangel.ajanovski@finki.ukim.mk

Mirela Gutica
British Columbia Institute of Technology
Burnaby, Canada
mirela_gutica@bcit.ca

Timo Hynninen
South-Eastern Finland University of Applied Sciences
Mikkeli, Finland
timo.hynninen@xamk.fi

Antti Knutas
LUT University
Lappeenranta, Finland
antti.knutas@lut.fi

Juho Leinonen
University of Helsinki
Helsinki, Finland
juho.leinonen@helsinki.fi

Chris Messom
Monash University
Melbourne, Australia
christopher.messom@monash.edu

Soohyun Nam Liao
University of California San Diego
San Diego, USA
snam@eng.ucsd.edu

ABSTRACT
The ability to predict student performance in a course or program creates opportunities to improve educational outcomes. With effective performance prediction approaches, instructors can allocate resources and instruction more accurately. Research in this area seeks to identify features that can be used to make predictions, to identify algorithms that can improve predictions, and to quantify aspects of student performance. Moreover, research in predicting student performance seeks to determine interrelated features and to identify the underlying reasons why certain features work better than others. This working group report presents a systematic literature review of work in the area of predicting student performance. Our analysis shows a clearly increasing amount of research in this area, as well as an increasing variety of techniques used. At the same time, the review uncovered a number of issues with research quality that drives a need for the community to provide more detailed reporting of methods and results and to increase efforts to validate and replicate work.

CCS CONCEPTS
- Social and professional topics → Computer science education.

KEYWORDS
educational data mining, analytics, learning analytics, prediction, performance, literature review, mapping study

ACM Reference Format:

1 INTRODUCTION
In the psychology and education-related scholarly literature, work on determining factors that contribute to academic performance has existed at least for a century. For example, in the late 1910s, a wide range of tests such as verbal memory tests were conducted with freshmen in an attempt to tease out factors that correlate with academic performance [45, 246, 397]. While the early work was conducted within psychology, interest in identifying individuals with particular performance characteristics such as the ability to program emerged soon thereafter [256].

While the earlier studies on identifying programming ability were mostly focused on attempting to find individuals who could perform as programmers [236], subsequent work spanned to trying to identify factors that predict students' computer aptitude [112] and performance in programming courses [340]. The reasons for predicting aptitude are numerous of which Evans and Sinkin outline several [112]: identifying potential majors, discriminating among applicants, advising students, identifying productive...
individuals, identifying those who may best profit from additional guidance, improving classes, determining importance of existing predictors, and exploring the relationship between programming abilities and other cognitive reasoning processes.

Parallel to this, there has been a stream of research that studies why students struggle with learning to program. A rather famous example of this work is the research on the rainfall problem, where students are expected to write a program that reads in numbers and then calculates their average [347]. While the problem has been studied rather extensively [333], one of the interesting approaches taken to study the problem was the instrumentation of working environments. For example, when studying students working on the rainfall problem, Davies employed a system that recorded key presses for further analysis [88].

Since then, systems that record students’ working process have been further adopted into computing education research [168]. This adoption of instrumented tools, combined with modern hardware, has created an opportunity to directly observe and react to student data, which has invigorated the research in models that can be used to predict academic performance.

This report is the outcome of an ITiCSE working group that is seeking to connect various communities – including those outside of computing education – that are supporting the work of predicting academic performance in computing courses. The working group is composed of internationally-diverse members with a variety of academic interests, and it worked for a period of three months, including an intensive five-day meeting at ITiCSE in July 2018.

We outline the results of a systematic literature review containing 357 articles that describes the breadth of work being done on the prediction of student performance in computing courses. In addition to the review itself, which summarizes the types of performance being predicted and the factors and methods used to perform the predictions, we identify trends in feature and method use over time, and offer insights obtained as we read. We believe this work is a mapping that will help to connect researchers in this area by identifying clusters of related work being published in different venues and highlighting opportunities for collaboration, integration, and broader dissemination.

1.1 Research Questions and Scope

The group initially sought to find literature related to identifying students who are “academically at-risk” [148], but initial forays into the literature highlighted that the term at-risk is often used to identify youth in disadvantaged circumstances. Several other terms, including performance, were used to focus the work on students at academic risk. As a result, the prediction of student performance became the focus of the group. We explored the following research questions:

(1) What is the current state of the art in predicting student performance?
   (a) How is performance defined? What types of metrics are used for describing student performance?
   (b) What are the features used for predicting performance?
   (c) What methods are used for predicting performance?
   (d) Which feature and method combinations are used to predict which types of student performance?
(2) What is the quality of the work on predicting student performance?

Both the metrics describing students performance and features used to predict them can be considered as variables. In the following, we will call input variables as features and output variables (i.e., performance metrics) as predicted (or output) variables. The term method refers to how output variables have been derived (possibly from the inputs).

The search terms used to identify the articles for this study are described in Section 3, but we include a discussion of the term performance here to define the scope of this work. For the purposes of this literature review, we defined the term broadly, including performance on a single assessment, performance in a course, retention in or continued progress through a program, and successful matriculation from a program. However, we only considered prediction of quantifiable metrics that are directly related to the course or the program such as grades, pass/fail probability, or retention in a program. We do not include articles that predict proxies, such as depression or team cohesion, that are not quantifiable or directly related to academic performance, even if they are likely to affect it. Similarly, we exclude articles that deal with predictions that may be a product of academic performance, such as employability, or articles that do not directly predict performance, such as those that primarily evaluate pedagogical interventions.

As a result of this broad focus on performance, our review covers a wide range of factors and contexts. As shown in Figure 1, students face challenges at many points in their academic career, and the factors contributing to those challenges vary widely, from family and economic stress to issues with academic preparation and a lack of study skills. We did not exclude any factors or methods that were identified in the articles that were included, though we attempted to cluster them into higher-level tags, such as “wealth” or “social factors” due to the diversity of individual factors examined.

1.2 Report Outline

This report is organized as follows. In the subsequent section, Section 2, we provide an overview of the existing review and survey articles on predicting student performance. The section discusses the exclusion and inclusion criterion, time spans, and suggested terminology and taxonomies in the existing literature. We draw upon this work when formulating our search and inclusion criteria, which are described in Section 3. Sections 4 and 5 present the results of the review, with Section 4 focusing on descriptive statistics and the results of a textual analysis and Section 5 presenting issues that emerged as reviewers completed the process. We provide several calls to the community and highlight possible directions for future work in Section 6. The list the articles reviewed is provided in Table 13, and we provide the form used to extract data from each reviewed article in an appendix. The form is provided so that readers can review our method, identify sources of bias, and if they wish, modify or extend our form for their use in similar projects.

2 SYNTHESIZING PREVIOUS REVIEWS ON PREDICTING STUDENT PERFORMANCE

This section provides a synthesis of previous reviews and surveys on predicting student performance. Our goal was to identify categories and trends already identified in existing reviews, so that our review uses common terminology, where possible, while also contributing a new understanding of the literature in the field. The synthesis described in this section informed the instrument we used in our study. A description of the data extraction instrument is provided in Section 3.1.3, and details are provided in Appendix A.

The corpus for the synthesis was created through metadata searches on the article indexes Scopus, IEEE Xplore, and the ACM Digital Library. For the search, the search string “student AND predict* AND performance AND (review OR survey)” – or the equivalent variant expected by each article engine – was used. The search was conducted in June 2018, and resulted in a total of 147 hits, including duplicates that were indexed in multiple sites. Manual inspection of article titles and abstracts and exclusion of one-page extended abstracts reduced the count to 13 reviews and surveys of the area. Because of the focus on performance prediction, broad reviews of educational data mining and learning analytics, such as [39, 102, 168, 269, 318], were omitted.

The quality of the reviews varied. In particular, the bounds of the search performed in the review or survey were not always well defined. Of the 13 articles, eight [160, 185, 198, 204, 252, 253, 269, 335] listed at least some of
Academic Loss Points: Why Do Students Drop?

Figure 1: Points in a student’s academic life, using CS as an example, where data for performance prediction can be gathered and performance prediction can be done. Both the first and second year loss factors can continue, to a lesser extent, through to graduation.

Table 1: High-level categorization of factors that can be used for predicting performance from the articles included in the synthesis.

<table>
<thead>
<tr>
<th>Study Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>[160] Activity and course features, Demographic features, Learning behavior features, Self-reported features, Student history record and performance, Student record and performance in current course, Others / unclear features</td>
</tr>
<tr>
<td>[198] Academic, Family, Institutional, Personal, Social</td>
</tr>
<tr>
<td>[204] Academic performance, Socio-economic, Personal information</td>
</tr>
<tr>
<td>[253] Learning behaviour data, Learning network data, Learning level data, Learning emotional data, Other</td>
</tr>
<tr>
<td>[335] Cumulative Grade Point Average, Engage time, External assessments, Extra-curricular activities, Family support, High-school background, Internal Assessment, Social interaction network, Study behavior, Student demographic, Student interest</td>
</tr>
</tbody>
</table>

the sources, venues or fields that they had searched. Only six [198, 204, 252, 253, 269, 335] listed the keywords or search terms used. Eight [99, 160, 185, 198, 252, 253, 269, 335] described a year range or the absence thereof. All surveyed a limited number or range of articles.

In this review of surveys, we collected and synthesized what these existing reviews and surveys reported about factors for predicting student performance, methods used for predicting student performance, and how the articles described performance. Of the reviewed articles, five [160, 198, 204, 253, 335] summarized factors, two [252, 335] summarized methodologies, and four [160, 198, 204, 252] offered insights into the meaning of performance or what was being predicted. None of the surveys discussed all three of these issues, but five [160, 198, 204, 252, 335] discuss two of them.

2.1 Factors

Table 1 contains a listing of the high-level factors used for predicting student performance. The category sizes range widely, from 3 to 10, indicating that the review authors saw value in various levels of abstraction when categorizing factors. We drew inspiration from this list when constructing the form we used to extract data. Our aim was to create a broad categorization that enables identification of the source of the data being used, so we synthesized the categories used in the reviews into Family Background, Demographic Data, Working Conditions, Educational Background, Course Data (current or parallel), Student Motivation, and Psychological / Affective / Learning Scales. We split motivation and other scales into two categories because motivation is a complex construct that has been extensively used and studied for predicting performance, so we expected to see a variety of types of motivation (and corresponding instruments) used in the articles we reviewed.

2.2 Methods

Muthukrishnan et al. [252] categorized methods used for predicting performance into four high-level categories: Decision trees, Regression, Clustering, Dimensionality reduction / other. Overall, other articles – if they included methods for predicting performance – did not explicitly attempt to provide a categorization, but provided lists of methods that have been used for prediction. In our form, we used the high-level categories Classification (supervised learning), Clustering (unsupervised learning), Mining (finding frequent patterns and/or feature extraction), and Statistical (correlation, regression, t-testing, etc), and we encouraged reviewers to provide details on the specific techniques used underneath these high-level categories.

A few of the articles compared the performance of methods through metrics such as accuracy that were extracted from the surveyed literature. While this approach has appeal, and we did extract such performance metrics from the articles we reviewed, we found that the contexts of the studies varied so significantly and the quality of reporting varied so widely that a meta-review would not be meaningful. A comparison of reported accuracy ratings would, in our opinion, lead the reader to misconceptions about the performance of various methods.

2.3 Definitions of Performance

Most of the articles included in the synthesis did not explicitly define performance. This may be a consequence of the articles being reviews of other
articles, effectively creating a situation where other articles are included as
long as they predict performance – no matter if they defined what it means.
The closest to a definition of performance was provided in [160], where
the authors suggest that, "studies tended to predict course performance
(successful/unsuccessful), course grades and student retention/dropout in
online/blended learning contexts."

We found the lack of definition problematic, as it led some surveys to
include articles with significantly different goals. Therefore, we reviewed
the articles for trends to see what was most included. Many agreed that
assessments, defined broadly, are the key metric. For example, one review
notes that:

... students performance can be obtained by measuring the
learning assessment and co-curriculum. However, most of
the studies mentioned about graduation being the measure
of students success. Generally, most of higher learning insti-
tutions in Malaysia used the final grades to evaluate students
performance. Final grades are based on course structure, as-
essment mark, final exam score and also extracurricular
activities. [335]

Another article, from 2017, agrees that student performance can be
observed using internal assessment metrics:

Most of the Indian institution and universities using [sic] fi-
nal examination grade of the student as the student academic
performance criteria. The final grades of any student depend
on different attributes like internal assessment, external as-
essment, laboratory file work and viva-voce, sessional test.
The performance of the student depends upon how many grades a student score in the final examination. [198]

However, some articles define performance more broadly. For exam-
ple, [204] defines performance as, "... a measure of a student’s competence
for future courses." They note that:

In addition to passing or failing a course, the grade obtained
is also of interest to course instructors. Furthermore, aca-
demic advisors would be interested in the time required for a
student to complete a degree and his/her ability of enrolling
in multiple programs. [204]

We agree with the inclusion of these wider factors. As a result, our data
extraction form initially included a range of academic performance factors,
including various forms of internal course assessment, but also including
program retention and likelihood of graduation as possible values to be
predicted.

Another survey, which provided an overview of the values being pre-
predicted in the articles they reviewed, agrees that the literature defines per-
formance broadly. They noted that the majority of articles they saw predicted
final grades in courses, but they saw other predictions being made when the
context is outside of the traditional classroom [252]. This led us to include
several questions about the population being studied in our data extraction
form, including the topic of the course in which the prediction is being
employed, the format of the course (e.g., traditional, online, hybrid, MOOC,

3 SYSTEMATIC LITERATURE REVIEW

In this section, we will first describe the how the literature review was
conducted (Section 3.1) and then relate statistics to describe the included
articles (Section 3.2). We consider our work to be a Systematic Literature
Review (SLR). However, the methodological distinction between systematic
mapping studies and SLR studies is often blurred [190, 283]. Our work can be
argued to be somewhere between these two approaches. In particular, we
aim to create a high level synthesis but we will not describe each identified
work separately as some SLR studies do.

3.1 Methodology

3.1.1 Identification of Relevant Literature. We began by collecting articles
about predicting student performance that were already known to the
experts in the working group. Based on this set of known articles, we
tested multiple search terms by looking at the following three indexes: (1)
Scopus, (2) IEEE, and (3) ACM. We started with the search terms used in
the previous surveys (see Section 2). After multiple iterations, we decided
to use the following search string:

(at-risk OR retention OR persistence OR attrition OR performance)
AND
(prediction OR modelling OR modeling OR detection OR predict OR
"machine learning") AND
("computer science" OR informatics OR engineering OR program-
ming OR cs)

The searches were conducted in June 2018. The syntax of the search
strings was adjusted for each index. After combining the results, filtering out
articles not written in English, and using an automated process to remove
duplicates based on article title, publication year, and DOI, a corpus of 4,200
articles was created. The working group manually reviewed the article
titles and abstracts to flag potentially relevant articles (see Section 3.1.2
for details). A total of 743 articles were flagged. We removed all articles
published prior to 2010 from this list, resulting in a set of 565 articles for
close analysis. (2010 was selected as a starting point because we detected a
marked increase in articles published at this point; it also provides a focus
on recent work.) At this point, only the articles clearly out of our scope or
published prior to our cutoff had been removed.

Review of the included articles identified that several known to the group
were not found in the set to be reviewed. In most cases, it was determined
that the articles were not included because they did not focus on prediction
of performance – as defined in the previous section. Instead, they focused
on modeling learners’ skills or on the discussion of factors that might be
related to performance, but without clear predictive goal. After consultation,
we decided to omit those articles. In other cases, we found that the article’s
publication venue was not indexed in any of our sources. In particular,
articles from the International Conference of Educational Data Mining were
not included. We manually added the relevant articles from this source to our
list, leading to a final total of 889 articles for analysis.

3.1.2 Inclusion Criteria. The inclusion criteria was the article must dis-
cuss predicting student academic performance. This necessitated a working
definition of academic performance, which we developed in the previous
section. In particular, as we flagged articles for review, we looked for a
value to be predicted that was related to a single assessment, a course, or
progress through a program. We only considered quantifiable metrics that
are directly related to a course or program that students are enrolled in,
such as course activities, exercise points, etc. We did not include articles
that predict proxies not directly related to academic performance of an
individual. More explicitly, work was not included if its focus was:

• Predicting team performance and dynamics
• Work placement
• Affective states (e.g., happiness, anxiety, depression)
• Intent (e.g., intent to enter a university)
• Automatic assessment tools (e.g., automatically assessing programming problems, automatically assessing essays, detecting plagiarism, teaching interventions, and recommender algorithms), if the article did not clearly and separately discuss performance prediction in an included context.

In rare cases, authors stated that a journal publication was constructed from a conference publication and included all the relevant information from the previous version of the article. In these cases, only the journal version was included. Similarly, theses and dissertations were removed, as related journal articles were included. This resulted in 497 articles remaining for a more detailed review.

3.1.3 Data Extraction. Based on the research questions and the meta-survey presented in Section 2, a preliminary version of the data extraction form was constructed. The working group members were divided into pairs, and each pair evaluated five articles using the form. Afterwards, the form was adjusted based on their responses to allow reviewers to provide more precise details about factors and methods. The high-level taxonomy of the resulting instrument is as follows:

- Prediction
  - What is being predicted
  - Details on what is being predicted
- Context
  - Number of subjects
  - Population
  - Course topic
  - Training mode
  - Education type
- Data features used to predict performance
  - Data set information
  - Family background (parental information, socioeconomic status, etc.)
  - Demographic data
  - Working Conditions
  - Education (background)
  - Course data (current or parallel)
  - Student motivation
  - Psychological / affective / learning scales
  - Used surveys / standardized questionnaires
- Methods and techniques
  - Method type
  - Classification (supervised learning)
  - Clustering (unsupervised learning)
  - Mining (finding frequent patterns / feature extraction)
  - Statistical techniques
  - Results details
- Quality factors of the article
  - Is there a clearly defined research question?
  - Is the research process clearly described?
  - Are the results presented with sufficient detail?
  - Does the article discuss threats to validity?
  - Are there separate training and prediction data sets (where relevant)?
  - Has the work been verified in a second population?
  - Are the data collection instruments linked or included?
  - Were all the features described in sufficient detail to identify them?
  - Additional notes on quality

To avoid ambiguity in the data extraction, all categories were implemented as check-boxes with an additional open-text field at the end for listing features not in the predefined list of options. Details of the form are provided in Appendix A, where all of the options provided are listed.

We provide these details partially to allow review of our data collection instrument but also in the hope that others may find the instrument useful.

Despite our attempts to list all of the common answers, the open text boxes were frequently used. In some cases, it even had to be used to document multiple items, so we established a protocol of separating items with a semicolon. Our parsing scripts extracted all of the items identified, resulting in lists of items for each high-level item (context, data features, methods, and quality factors). Pairs of reviewers reviewed each list to partition related items. For example, all statistical tests of variance were collected into a single category, and this category was used in the reports in Section 4.

3.2 Descriptive Results

During data extraction, reviewers continued to apply the inclusion criteria and excluded articles that did not explicitly discuss prediction of academic performance. In total, data was obtained from 357 articles. Table 2 presents the number of publications per year. Data from 2018 should be omitted from an analysis of long-term trends, since the work was completed in July, leaving half a year for additional work to be published. Focusing on 2010 through 2017, then, we see an unmistakable increase in work published each year.

Table 2: Number of papers reviewed per year of publication.

<table>
<thead>
<tr>
<th>Years</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>7</td>
</tr>
<tr>
<td>2011</td>
<td>20</td>
</tr>
<tr>
<td>2012</td>
<td>22</td>
</tr>
<tr>
<td>2013</td>
<td>36</td>
</tr>
<tr>
<td>2014</td>
<td>43</td>
</tr>
<tr>
<td>2015</td>
<td>53</td>
</tr>
<tr>
<td>2016</td>
<td>70</td>
</tr>
<tr>
<td>2017</td>
<td>75</td>
</tr>
<tr>
<td>2018</td>
<td>31</td>
</tr>
<tr>
<td>Total</td>
<td>357</td>
</tr>
</tbody>
</table>

Table 3 lists the disciplines in which prediction was performed. Recall that our search terms focused on computer science, engineering, and informatics, so it’s unsurprising that most of the work we reviewed performed prediction in a CS or, more generally, STEM context. Mathematics consists of one third of STEM. The “Multi disciplinary” category refers to work that explicitly predicted performance in two or more contexts in different disciplines. Most of these were within CS or STEM as well. The rationale of this preliminary analysis is to illustrate the scope of this survey: the focus truly is in engineering and mostly in computing.

Table 3: The discipline in which the prediction was being performed, if a specific discipline is named.

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>126</td>
<td>34.9%</td>
</tr>
<tr>
<td>STEM</td>
<td>98</td>
<td>27.1%</td>
</tr>
<tr>
<td>Other</td>
<td>39</td>
<td>10.8%</td>
</tr>
<tr>
<td>Multi disciplinary</td>
<td>30</td>
<td>8.3%</td>
</tr>
<tr>
<td>Unclear</td>
<td>14</td>
<td>3.9%</td>
</tr>
</tbody>
</table>

Table 4 presents the venues in which work has been published. We only list venues with three or more reviewed articles due to the large number of specialized venues which contributed one or two articles. In the venues with multiple articles, we saw computing education (SIGCSE, ITiCSE, ICER, etc.), engineering education (FIE, ICEED, EDUCON, etc.), STEM education (ISEC), and learning analytics and data mining (LAK, EDM, L@S, etc.).
About half of the articles we reviewed (171) were the only work from a particular venue, suggesting that there is broad interest in the topic. The single list of single-article venues includes conferences focusing on software engineering, machine learning, general education, and psychology. The topic is of interest to researchers and practitioners in a wide variety of fields, so we find work tuned to specific disciplines and regions in venues catering to those communities. Unfortunately, this result also suggests that the community is highly dispersed, which makes dissemination, collaboration, and validation challenging.

3.2.1 Topic modeling and discovered themes. As a part of our exploratory mapping, we sorted the documents into topics using the Latent Dirichlet Allocation (LDA) algorithm [51]. We used LDA as a statistical text mining method for assigning documents into topics, which are detected using word association and distributions [50]. The underlying mechanism in LDA is a probabilistic Bayesian network model, in which each document is characterized by certain topics, and each topic is defined by a specific set of words, which co-occur with a certain probability. To summarize, the topics word association and distributions [50]. The underlying mechanism in LDA is a probabilistic Bayesian network model, in which each document is characterized by certain topics, and each topic is defined by a specific set of words, which co-occur with a certain probability. To summarize, the topics word association and distributions [50]. The underlying mechanism in LDA is a probabilistic Bayesian network model, in which each document is characterized by certain topics, and each topic is defined by a specific set of words, which co-occur with a certain probability. To summarize, the topics word association and distributions [50]. The underlying mechanism in LDA is a probabilistic Bayesian network model, in which each document is characterized by certain topics, and each topic is defined by a specific set of words, which co-occur with a certain probability. To summarize, the topics word association and distributions [50].

For the analysis, we used a modified version of the `topicmodels` R package [155] and visualized with the LDAvis library [338]. Semantic coherence, a quality value for deciding the number of topic models [243], was calculated using the `R stm` library [313]. Additionally, the LDAvis library was also used to calculate the distance between topics on a scatter-plot, which approximates the semantic relationships between the topics with multidimensional scaling. It is a method similar to factor analysis and allows the level of similarity between objects to be visualized. The inter-topic distance was calculated using Jensen-Shannon divergence [338]. LDA-based topic modeling is a commonly used method for text analysis and equivalent methods have been used to statistically analyze scientific texts in previous studies [84, 180, 282, 387].

After examining the local maximums in the semantic coherence results, we proceeded with a topic model with 11 topics. The topic modeling results and the themes we discovered are summarized in Table 5. Topics whose relative size was less than two percent were excluded from analysis to avoid unjustified generalizations. The sum of the sizes adds up to just over 100% due to rounding issues.

All of the themes identified are connected due to their focus on students, prediction, and performance. In order to avoid repetition, these focuses are not explicitly mentioned in each row of the table. Furthermore, it should be noted that this initial exploratory categorization is based on probabilistic modeling and document corpora. The inter-topic distances and the prevalence of each topic are presented in Figure 2. The results are presented on two axes via multidimensional scaling, with prevalence of each topic denoted by size of the circle and distances by distances between the circles.

4 THEMATIC ANALYSIS

4.1 Predicted Values (Outputs)

To answer research question (1a), about the types of metrics are used for describing student performance, we categorized the types of performance that the studies aim to predict. Table 6 describes these values to predict the articles that were reviewed. Some articles predicted more than one category of value, so the percentages reflect the relative number of articles that showed interest in that method for describing student performance. The most popular indicator for performance is course grade, which is used as a predictor in a quarter of the studies examined, as well as an additional 13.6% that predicted a grade range. Overall, we can see that summative performance metrics, such as GPA, are the preferred prediction target. Other approaches, such as post-course outcomes or assessment of learning outcomes, are represented but are in the minority. Unfortunately, in 12.2% of the studies, it was not clear what the actual metric to predict student performance was, and therefore these metrics were classified as “Unspecified or Vague Performance.”

Table 7 lists the frequency at which various indicators of performance have been predicted over the years. As earlier, data from 2018 should not be considered when evaluating long-term trends. Most factors have experienced comparable growth over the entire period we observed, but prediction of “knowledge gain” has only appeared recently.

4.2 Features Used (Inputs)

Table 8 lists the frequency at which various features have been used to predict student performance. 29 input features were identified, with performance in the course of interest, engagement in the course of interest, and performance in previous courses standing out as the most common data being used in predictions.

General demographic data has been less frequently used recently, but some aspects of demographic data, like gender and health, have become increasingly important. There also appears to have been a surge in interest in some psychometric factors, like self-regulation and self-efficacy, and in data extracted from logs, such as time on task.

We also examined how many factors were used to perform a prediction, on average, and were disappointed to find that most articles used a small number of factors. While studies may appear to use multiple factors, in many cases, these factors are measuring similar things. There are few studies that combine different data sources, such as performance in previous courses, demographic data, and psychometric data. The variance was actually larger in earlier years of the study, suggesting that there may have been more exploratory articles in the early 2010’s, with current efforts focusing on the impact of a small number of factors.

4.3 Methods Used

Table 9 lists the frequency at which various statistical and machine learning methods have been used to predict student performance.
Table 4: Publication venues with at least three included papers.

<table>
<thead>
<tr>
<th>Venue</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Conference on Learning Analytics and Knowledge (LAK)</td>
<td>21</td>
</tr>
<tr>
<td>International Conference on Educational Data Mining</td>
<td>17</td>
</tr>
<tr>
<td>Frontiers in Education Conference (FIE)</td>
<td>13</td>
</tr>
<tr>
<td>ACM Technical Symposium on Computer Science Education (SIGCSE)</td>
<td>10</td>
</tr>
<tr>
<td>ASEE Annual Conference and Exposition</td>
<td>8</td>
</tr>
<tr>
<td>International Journal of Engineering Education</td>
<td>8</td>
</tr>
<tr>
<td>International Conference on Advanced Learning Technologies (ICALT)</td>
<td>7</td>
</tr>
<tr>
<td>ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE)</td>
<td>6</td>
</tr>
<tr>
<td>International Computing Education Research Conference (ICER)</td>
<td>5</td>
</tr>
<tr>
<td>International Conference on User Modeling, Adaptation, and Personalization</td>
<td>5</td>
</tr>
<tr>
<td>International Conference on Inventive Computation Technologies (ICICT)</td>
<td>4</td>
</tr>
<tr>
<td>IEEE International Conference on Engineering Education (ICEED)</td>
<td>4</td>
</tr>
<tr>
<td>IEEE Global Engineering Education Conference (EDUCON)</td>
<td>4</td>
</tr>
<tr>
<td>Integrated STEM Education Conference (ISEC)</td>
<td>4</td>
</tr>
<tr>
<td>Computers &amp; Education</td>
<td>4</td>
</tr>
<tr>
<td>IEEE Transactions on Learning Technologies</td>
<td>3</td>
</tr>
<tr>
<td>Annual Conference on Information Technology Education (SIGITE)</td>
<td>3</td>
</tr>
<tr>
<td>IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT)</td>
<td>3</td>
</tr>
<tr>
<td>Journal of College Student Retention: Research, Theory &amp; Practice</td>
<td>3</td>
</tr>
<tr>
<td>ACM Conference on Learning @ Scale</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5: Summary of discovered topic modeling-based themes

<table>
<thead>
<tr>
<th>Topic</th>
<th>Theme</th>
<th>Relative size</th>
<th>Papers in the theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1 (excluded because of small size)</td>
<td>0.29%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 2 (excluded because of small size)</td>
<td>0.58%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 5 (excluded because of small size)</td>
<td>0.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 7 (excluded because of small size)</td>
<td>0.58%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 9 (excluded because of small size)</td>
<td>3.2%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistical methods such as linear modeling (31.3%) including linear regression and ANOVA are the most common methods we observed, with various graph models and decision trees close behind. In general, classification techniques are used more frequently than clustering techniques (and our reviewers note that clustering is occasionally used as a preparatory step before applying a model). The variety of unique methods used has increased in the period we observed.

4.4 Cross-tabulation of Features (Inputs) to Performance Values (Outputs)

Finally, we explored the relationship between the value being predicted and the inputs used in the prediction. Figure 3 is an association map that illustrates the most frequently observed combinations of inputs and outputs. For example, course grades are often predicted by other grades and
Table 6: Values being predicted in the reviewed papers. Some papers predicted more than one value.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Grade or Score</td>
<td>88</td>
<td>24.4%</td>
</tr>
<tr>
<td>Exam / Post-test Grade or Score</td>
<td>53</td>
<td>14.7%</td>
</tr>
<tr>
<td>Course Grade Range (e.g., A-B/C-F, Pass/Fail)</td>
<td>49</td>
<td>13.6%</td>
</tr>
<tr>
<td>Program or Module Graduation / Retention</td>
<td>48</td>
<td>13.4%</td>
</tr>
<tr>
<td>Unspecified or Vague Performance</td>
<td>44</td>
<td>12.2%</td>
</tr>
<tr>
<td>GPA or GPA Range (including CGPA, SGPA)</td>
<td>44</td>
<td>12.2%</td>
</tr>
<tr>
<td>Assignment Performance</td>
<td>41</td>
<td>11.4%</td>
</tr>
<tr>
<td>Course Retention / Dropout</td>
<td>20</td>
<td>5.5%</td>
</tr>
<tr>
<td>Knowledge Gain</td>
<td>8</td>
<td>2.2%</td>
</tr>
<tr>
<td>Number of Courses Passed or Failed</td>
<td>4</td>
<td>1.1%</td>
</tr>
</tbody>
</table>

Figure 3: Most frequently researched features as predictors (left side) for predicted values (right side). The graph only includes links that were explored by at least 10 articles. The thickness of the links denotes the number of articles that explored such a predictor.

5 DISCUSSION

In the previous section, we presented a description of the data we obtained from reviewing the articles identified by our review. In this section, we offer insights into the data that may not be explicitly reflected in the data but which we experienced as we completed the review.

5.1 Publication Trends

Tables 8 and 9 contain the number of times that factors and methods, respectively, were used in articles in various years. While the unique number of methods and factors grew in the first years we observed, it has not obviously increased over the past four years. However, we did see many recent articles experimenting with variants of previously used methods. We also saw increasing numbers of articles using machine learning techniques, such as support vector machines (SVMs) and probabilistic graph models (such as Hidden Markov Models). Similarly, while the number of unique high-level factor categories has not increased in recent years, we saw a shift in the particular factors being used, with submissions, log data, and similar artifacts being much more heavily utilized in recent years.

As we noted earlier, we did not see as many demographic factors being used to predict course grades, and that may be an area for future investigation. We also saw a general absence of work that used motivation, which we had expected to see. Instead, many articles seemed to be using engagement, perhaps as a proxy. More generally, psychometric data is less frequently utilized, and the increase in usage of self-efficacy and self-regulation data might be signalling growth in that area.

5.2 Contexts

Most of the work we reviewed was performed in a post-secondary context, and the figures in the previous section should generally be interpreted as applying to that educational level. However, we did observe some work in the K-12 and secondary environments. GPA, single course, and program retention predictions are not (or are less) relevant in these environments, and much of the work we observed was predicting interest (which we excluded from this study) or performance in modules or individual exercises.

We also briefly considered the relationship between the discipline being studied and the value being predicted. This analysis is presented in Table 11. It appears that the disciplines are being investigated in a fairly uniform manner. However, as discussed earlier, most of the work identified here is in the context of engineering. Findings in other disciplines might be different.

5.3 Quality

During our review, we evaluated articles on several aspects of quality. Table 12 displays the results of this effort. Several of the results are disheartening. For example, in almost one out of ten articles, we had trouble identifying what was being predicted (e.g., "student performance" or "student success" without qualification).

5.3.1 Reporting Results. Other attributes are also directly related to our ability to interpret the results. Only a third (33%) of the articles we examined included a directly stated research question. Another 40% stated their goals without presenting them as research questions. In some cases, we also had trouble identifying how the prediction was being made and whether the data was reliable. In several articles, it was difficult to determine which data was being used to perform the prediction. For example, data presented in the methodology section might have been used for prediction or simply presented as demographic data to help describe the population. More seriously, in about a third of the articles we reviewed, the features being used in prediction were not described in sufficient detail for us to identify them with confidence. For example, some articles indicated that "interest," "personality traits," "motivation," or "stress" were features, but these terms can reflect different quantities. Standard scales for measuring some of these features...
### Table 7: The aspects of student performance predicted over the years

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Grade or Score</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>14</td>
<td>12</td>
<td>15</td>
<td>24</td>
<td>7</td>
<td>84</td>
<td>22.28%</td>
</tr>
<tr>
<td>Exam / Post-test Grade or Score</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>7</td>
<td>51</td>
<td>13.53%</td>
</tr>
<tr>
<td>Course Grade Range (e.g. A–B/C–F, Pass/Fail)</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>47</td>
<td>12.47%</td>
</tr>
<tr>
<td>Program or Module Graduation / Retention / Dropout</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>12</td>
<td>7</td>
<td>44</td>
<td>11.67%</td>
</tr>
<tr>
<td>GPA or GPA Range (incl. CGPA, SGPA)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>41</td>
<td>10.88%</td>
</tr>
<tr>
<td>Unspecified or Vague Performance</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>13</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>40</td>
<td>10.61%</td>
</tr>
<tr>
<td>Assignment Performance (e.g. grade, time to completion)</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>40</td>
<td>10.61%</td>
</tr>
<tr>
<td>Early Detection of Course Retention / Dropout</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>18</td>
<td>4.77%</td>
</tr>
<tr>
<td>Knowledge Gain</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>1.06%</td>
</tr>
<tr>
<td>Number of Courses Passed or Failed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 8: Use of factors to predict student performance by year

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Performance</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>10</td>
<td>23</td>
<td>23</td>
<td>26</td>
<td>34</td>
<td>12</td>
<td>141</td>
<td>13.09%</td>
</tr>
<tr>
<td>Pre–Course Performance</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>13</td>
<td>16</td>
<td>23</td>
<td>21</td>
<td>33</td>
<td>15</td>
<td>139</td>
<td>12.91%</td>
</tr>
<tr>
<td>Engagement</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>17</td>
<td>31</td>
<td>27</td>
<td>5</td>
<td>113</td>
<td>10.49%</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>19</td>
<td>14</td>
<td>12</td>
<td>16</td>
<td>10</td>
<td>86</td>
<td>7.99%</td>
</tr>
<tr>
<td>Personality</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>14</td>
<td>14</td>
<td>6</td>
<td>65</td>
<td>6.04%</td>
</tr>
<tr>
<td>Demographic</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>14</td>
<td>13</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>65</td>
<td>6.04%</td>
</tr>
<tr>
<td>Secondary School Performance</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>11</td>
<td>12</td>
<td>5</td>
<td>11</td>
<td>5</td>
<td>58</td>
<td>5.39%</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>11</td>
<td>7</td>
<td>53</td>
<td>4.92%</td>
</tr>
<tr>
<td>Family</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>52</td>
<td>4.83%</td>
<td></td>
</tr>
<tr>
<td>Task Time</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>11</td>
<td>9</td>
<td>5</td>
<td>41</td>
<td>3.81%</td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>33</td>
<td>3.06%</td>
<td></td>
</tr>
<tr>
<td>Self–Regulation</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>2.60%</td>
<td></td>
</tr>
<tr>
<td>Log Data</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>28</td>
<td>2.60%</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>2.32%</td>
<td></td>
</tr>
<tr>
<td>Discipline</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>1.86%</td>
<td></td>
</tr>
<tr>
<td>Wealth</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>19</td>
<td>1.76%</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>19</td>
<td>1.76%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>19</td>
<td>1.76%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School Quality</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>19</td>
<td>1.76%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self–Efficacy</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>13</td>
<td>1.21%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post–Course Performance</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>13</td>
<td>1.21%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>11</td>
<td>1.02%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td>0.74%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concept Assessment</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0.56%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Browsing History</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0.28%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

183
Table 9: Use of methods to predict student performance by year. Only methods identified at least five times are included.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical: Linear Modeling</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>9</td>
<td>9</td>
<td>14</td>
<td>27</td>
<td>12</td>
<td>110</td>
<td>17.71%</td>
</tr>
<tr>
<td>Classification: Probabilistic Graphical Model</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>12</td>
<td>11</td>
<td>13</td>
<td>18</td>
<td>70</td>
<td>12.88%</td>
</tr>
<tr>
<td>Classification: Decision Trees</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>12</td>
<td>11</td>
<td>16</td>
<td>16</td>
<td>74</td>
<td>11.92%</td>
</tr>
<tr>
<td>Statistical: Correlation</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>57</td>
<td>9.18%</td>
</tr>
<tr>
<td>Classification: Neural Network</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>11</td>
<td>51</td>
<td>8.21%</td>
</tr>
<tr>
<td>Classification: SVM</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>11</td>
<td>5</td>
<td>45</td>
<td>7.25%</td>
</tr>
<tr>
<td>Classification: Classification</td>
<td>3</td>
<td>5</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>3</td>
<td>42</td>
<td>6.76%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical: Latent Variable Models</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>27</td>
<td>4.35%</td>
<td></td>
</tr>
<tr>
<td>Classification: Random Forest</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>25</td>
<td>4.03%</td>
</tr>
<tr>
<td>Clustering: Partitioning-Based</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>19</td>
<td>3.06%</td>
<td></td>
</tr>
<tr>
<td>Classification: Nearest Neighbor</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>17</td>
<td>2.74%</td>
</tr>
<tr>
<td>Statistical: Variance Analysis</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>13</td>
<td>2.09%</td>
</tr>
<tr>
<td>Classification: Smote</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>10</td>
<td>1.61%</td>
</tr>
<tr>
<td>Mining: Text</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>1.45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classification: Logistic Regression</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>1.45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical: PCA</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>1.13%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mining: Time</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0.81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mining: Mining</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0.81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mining: Association Rule</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0.81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustering: Hierarchical Clustering</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0.81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustering: Density-Based</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0.81%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10: Cross-tabulation of features (inputs) and performance values (outputs).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Grade or Score</td>
<td>42</td>
<td>35</td>
<td>35</td>
<td>19</td>
<td>10</td>
<td>12</td>
<td>18</td>
<td>10</td>
<td>21%</td>
</tr>
<tr>
<td>Program or Module Graduation / Retention / Dropout</td>
<td>11</td>
<td>28</td>
<td>5</td>
<td>25</td>
<td>24</td>
<td>20</td>
<td>14</td>
<td>15</td>
<td>13%</td>
</tr>
<tr>
<td>Course Grade Range (e.g. A–B/C–F, Pass/Fail)</td>
<td>20</td>
<td>22</td>
<td>19</td>
<td>15</td>
<td>11</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>12%</td>
</tr>
<tr>
<td>Exam / Post-test Grade or Score</td>
<td>32</td>
<td>17</td>
<td>23</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>9%</td>
</tr>
<tr>
<td>Unspecified or Vague Performance</td>
<td>15</td>
<td>18</td>
<td>11</td>
<td>13</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>12%</td>
</tr>
<tr>
<td>GPA or GPA Range (incl. CGPA, SGPA)</td>
<td>9</td>
<td>18</td>
<td>3</td>
<td>13</td>
<td>8</td>
<td>15</td>
<td>15</td>
<td>7</td>
<td>15%</td>
</tr>
<tr>
<td>Assignment Performance (e.g. grade, time to completion)</td>
<td>20</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Early Detection of Course Retention / Dropout</td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>16%</td>
</tr>
<tr>
<td>Knowledge Gain</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6%</td>
</tr>
<tr>
<td>Number of Courses Passed or Failed</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>16%</td>
</tr>
</tbody>
</table>
The field of predicting student performance is large and so far the work has not converged sufficiently for us to be able to determine the state of the art. Instead, based on observations from the survey process, we highlight a set of dimensions in which research is being conducted; methodology, data, usability, and application.

The methodological dimension is a time-wise (and consequently partially hype-wise) continuum of methods used for data analysis. It ranges from basic statistical analysis and correlations to the use of state of the art machine learning methods. For example, as deep learning has recently received increased attention in the machine learning research community, it has also been applied in predicting student performance.

The data dimension consists of three parts: data collection, quantity and granularity, each of which form their own dimensions. First, there is a shift from using data collection methods that need to be manually extracted, such as surveys filled using pen and paper, to using data collection methods that are automatic, such as working environment instrumentation and online surveys. Second, there is a movement from using single-semester single-course data, typically with post-hoc analysis, towards data from multiple semesters or multiple courses. Third, there is a movement from using coarse-grained data such as course grades towards more fine-grained data such as instrumented working environment data and physiological measurements. These dimensions are also related to starting to take a more holistic view of a student; recently, for example, student health, self-regulation, and study skills have gained attention.

The usability dimension is related to the data quantity dimension: given the movement from single-course single-semester data towards data from multiple courses or semesters, we are seeing predictive models that are evaluated on separate data sets. This increase of variability in the data can lead to models that generalize better. This usability dimension is related to the final dimension, which is the application dimension. Here, we are moving from research that shows that something can be done to doing something with the data, for example within the context of intelligent tutoring systems.

These dimensions highlight the wide range of approaches in this area, and research in even one of the dimensions may lead to improved performance prediction models. At the same time, while researchers typically push for novel contributions, we must highlight the importance of replication studies – even seemingly simple change of context may lead to different results – and reattempting existing work where one of the dimensions is adjusted. For example, work that may have looked like a dead end some time ago could currently lead to new findings as research in the dimensions is evolving.
5.5 Comparing Approaches

Drawing on the above observations and multiple dimensions, we briefly discuss the meaningfulness of comparing approaches for predicting students' performance. Researchers who work on predicting performance typically consider at least the methodological and data dimensions. Here, the data comes from the context that is observed and in which future academic performance is predicted.

In our survey, we found almost no datasets that have been published for wider use, and furthermore, only approximately 25% of the reviewed articles linked or included the used data collection instruments. This reflects on the findings of the ITiCSE 2015 working group on educational data mining, who point out the scarceness of open data sources [168].

Researchers can compare methodological approaches on the data at their disposal. Similarly, if they have access to multiple datasets, they can compare methodological approaches on all of those data sets. At the same time, as long as the data that is used to make predictions is not made open or publicly available, or even shared to a smaller group of researchers, making justified comparisons across studies is not straightforward.

5.6 The Risk of Models

While everyone is enthusiastic about the opportunities to help students that point out the scarceness of open data sources [168], as "consent," "authorized," "permission," "IRB," "review board," or "ethics," members of the population being studied, or (c) with the explicit approval seeing very few explicit indications that the data used in the articles we examined if the articles we reviewed considered the risk of using certain methods to predict success has a high risk of causing harm. For example, if such a model were to be used for deciding who should enter a program, then the population would become increasingly privileged.

We have not, in this study, explicitly explored issues with bias or examined if the articles we reviewed considered the risk of using certain factors or publishing particular models. However, we saw little evidence that these issues are described in the literature we examined. We call on the community to be cautious in proposing and using models and ask that they consider the role that researchers and educators can play in public discourse.

5.7 Ethical Considerations

Finally, we raise several ethical issues resulting from our review: lack of consent, lack of inclusion, potentially unethical practices when collecting data, and issues of anonymity.

While we did not collect data on consent during the review, we recall seeing very few explicit indications that the data used in the articles we reviewed had been gathered (a) from public sources, (b) with consent from members of the population being studied, or (c) with the explicit approval of administrators responsible for institutional data. Textual analysis of the results confirms this impression. We searched the articles for indicators such as "consent," "authorized," "permission," "IRB," "review board," or "ethics," and found little evidence that approval for data collection is explicitly discussed. While not all institutions and societies require consent, we argue that it should at least be normal practice to acknowledge that ethical issues like the use of personal data, have been considered. We call for the community to normalize the reporting of the source of data and any consent or authorization processes (or the lack of a requirement for such processes) used in the collection of the data.

Very few of the articles we evaluated were explicitly inclusive with respect to gender. As far as we can tell, gender data was frequently presented with only two binary choices and without options like "another," "transgender," or "prefer not to disclose." It is also possible that a third option may have been present but was unused or unreported, but we could not, in many cases, evaluate this possibility since, in another example of reporting issues, the data collection instruments were not provided.

Data used in prediction studies is often secondary data as it is often gathered as normal practice during a course or program of study. Very few prediction studies recruit participants outside of an existing educational context or run formal experiments. Using secondary data can be problematic, as it may not have been anonymized appropriately and consent may not have been sought for use in a different context. For example, in one article, web traffic data from dormitories was monitored for normal network purposes, but the data was then stored and used in a study without notification.

Finally, we saw a few cases where participants were not adequately anonymized. This was, fortunately, rare, but in one notable example, actual student names appeared in a table alongside criteria used in student counselling.

We call on the reviewer community to screen for issues with anonymity and data collection and to request at the least the acknowledgement that ethical issues like consent have been considered and, where required, have been reviewed by an appropriate institutional review board.

5.8 Threats to Validity

As we were aware that systematic reviews have a number of limitations, we worked diligently to identify potential risks and to mitigate them.

A known risk of external validity is not having reviewed enough (or the appropriate) articles and including irrelevant material. The corpus for synthesis was created through metadata searches on the article indexes Scopus, IEEE Xplore, and ACM Digital Library. However, we found that some relevant articles are not indexed in these libraries. For example we discovered that articles published in the Journal of Educational Data Mining (JEDM) and LearnTechLib were not included in our corpus. We manually added articles from the Educational Data Mining community to our list, but it’s likely that other un-indexed venues were missed. Similarly, it’s possible that other search terms, such as the keyword “forecast,” would uncover more material. However, the literature we are aware of is included, and we believe we have included a large enough sample to be representative. With respect to the inclusion of irrelevant material, we were careful to define explicit inclusion criteria and practiced the application of these criteria in pairs before beginning data collection. We are also confident that we eliminated inappropriate entries, such as poster abstracts, during the manual inspection of article titles and abstracts.

The study design and consistent application of the review template are both potential internal validity issues. The risk related to study design is around the question: Did we miss anything? It is important to mention here that we extracted the data based on our research question. Looking to our data with a different research focus would result in other data being extracted. We have provided our extraction form in Section 3.1.3 for transparency and review. Another risk when using several reviewers to extract data is the lack of consistency. To mitigate this risk, we began by reviewing articles in pairs; only after this initial session we proceeded to individual extractions. Furthermore, the first two-thirds of the data collected was collected when we were working in the same physical location,
where questions were raised as they arose. The last third was collected asynchronously, but we feel that the reviewers were, by this point, comfortable with the extraction process.

6 CONCLUSION AND FUTURE WORK
The goal of this ITiCSE working group was to determine the current state of the research on predicting student academic performance. The work was conducted as a systematic literature review, reviewing relevant articles indexed by Scopus, IEEE Xplore and ACM indexes by June 2018. Works from the International Conference on Educational Data Mining were also examined, as the reviewers were aware of their existence and found that that the venue was not indexed in the databases searched. The final dataset that was analyzed contained 357 articles. These articles are listed in Table 13, categorized by the values they predicted.

6.1 Summarizing Research Question Results
Our main research questions of this work were: (1) What is the current state-of-the-art in predicting student performance? and (2) What is the quality of that work?

To summarize, during the recent years, there has been a clear increase in the amount of published research in the area, which can also be seen in the emergence of venues relevant for the topic. The majority of the work is looking at predicting easily attainable metrics such as individual course grade (38%), individual exam grade (14.7%), program retention or dropout (13.4%), GPA or cumulative GPA (12.2%), and assignment performance (11.4%). A small number of recent articles also examine measures that seek to better quantify learning performance, such as knowledge gain or speed in which the student will complete an assignment.

The features that have been used to predict student performance can be broadly split into five categories: demographic (e.g., age, gender), personality (e.g., self-efficacy, self-regulation), academic (e.g., high-school performance, course performance), behavioral (e.g., log data) and institutional (e.g., high-school quality, teaching approach). The majority of the articles used academic data for prediction (e.g., predicting course performance based on high-school performance). The use of data describing student behavior in a course (log data), while becoming more popular within the computing education research domain, is still relatively rare.

The methodologies that are used can be split into Classification (supervised learning, e.g., Naive Bayes, Decision Trees), Clustering (unsupervised learning, e.g., partitioning data), Statistical (e.g., correlation, regression), Data mining (identifying features and trends) and other methods. We found that (linear) regression models and classification methods are among the most frequently used tools, where the former is typically a method for the prediction, while for the latter the classification algorithms are often compared, leading to multiple prediction results.

When considering the quality of the existing work, there is room for improvement. The best articles that we read utilized data from multiple contexts and compared multiple methods to investigate a feature or variable of interest using multiple methods. However, we saw little re-use and sharing of data, which would allow us to compare methods or features, and we saw weaknesses in research methods and reporting. From the included articles, 33% included a clear research question, 18% discussed validity issues, and 8% verified the work in a second population. The last result echoes the finding of an ITiCSE working group on Educational Data mining [168], where the majority of the studies focused on a single course or a single institution, having no separate population with which the work would have been replicated with. On a positive note, 90% of the articles clearly defined what the target variable – or the predicted value – was, but this may partially be due to our inclusion criteria.

While there are no strong observable trends in emerging techniques, we highlighted a set of dimensions in which research is being conducted: methodology, data, usability, and application. Contributions to the body of predicting student performance can come in all of these dimensions – a researcher can, for example, study whether novel machine learning methods improve prediction when compared to other methods. There are areas with increasing interest such as the use of fine-grained log data, biometric data, and data describing students’ tendencies such as self-regulation questionnaires. So far, no silver bullet has emerged.

6.2 Calls to the Community
Based on our literature review (see Table 2), interest in this area is growing. However, to make future studies more meaningful and to enable researchers to build upon published results, our first call to the community is to improve reporting standards. As an example, we provide a checklist for a minimum viable article/paper (MVP) that is reporting on work predicting student performance. Table 14, provides a checklist that each article focusing on predicting student performance should include. The list is a minimum requirements list, which essentially outlines the need for explicit and clear methodology and results.

Our second call to the community echoes that of the ITiCSE 2015 Educational Data Mining working group. We call for open data sets for developing approaches for predicting student performance. Having open data sets or a baseline standard would help researchers compare their data and methods with the results of others, helping them direct their efforts to more accurate models.

Our third call to the community, as revealed by the number of venues discovered during the literature survey, is related to the highly distributed community. We call for explicitly seeking collaborators in other communities, perhaps as part of a replication or comparison effort – such an effort would also help disseminate advancements. Submitting to and attending venues where work from a particular project has not yet been published would also be welcome and effective for creating connections.

Our fourth call to the community is related to the lack of comparing and replicating existing work. So far, relatively little work replicates previous efforts and more, but still relatively few, articles explicitly compare published methods for predicting performance. Doing this work would help identify more effective methods and would also provide an opportunity for broader collaboration. Changes to the previous algorithms can be iterative by nature. For example, work that measures the impact of adding an additional data source to an existing approach or that uses underutilized data sources, such as psychometrics, would be interesting.

Our final call to the community is related to reporting outcomes. While there is a push towards developing new methods, data, and so on, publishing information on approaches that did not work is important. Some of this work is published implicitly, for example in research where feature selection methods rule out a set of variables. We hope, however, that researchers would highlight methods and features that did not work with a similar zeal that is used to report what worked. Only through understanding both what works and what does not can we form a holistic understanding of the topic.

A REVIEW EXTRACTION INSTRUMENT

- Initial Vocabulary/Taxonomy for Performance
  - Assignment grade
  - Course retention / dropout
  - Course grade
  - Course grade range (e.g. A-C, D-F)
  - Course pass / fail
  - Exam grade
  - GPA
  - Graduation
  - Program retention / dropout
  - Unspecified performance
  - Not applicable
  - Other
Describe threats to validity and limitations. Note situations in which a model or approach might be applied as well as where it is not valid.

Reflect upon the results and consider why the methods and factors used did work or did not work. What are the particular context-specific issues that may influence the outcomes?

Discuss the reasons why specific factors, performance metrics and methods were chosen (or omitted).

Provide the results. Perform and report on the tests necessary to test required attributes of the data. Name the analyses being performed and report all the relevant statistics to allow for interpretation of the results.

Discuss the relevant statistics to allow for interpretation of the results.

Define what is being predicted. If the value describing performance (e.g., course grade) consists of multiple items (e.g., course exam, course assignments), describe the contribution (weight) of each item when the performance value is calculated.

Table 13: The 357 reviewed papers organized by the value they predict

<table>
<thead>
<tr>
<th>Predicted Value</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Courses Passed or Failed</td>
<td>[120, 140, 176, 377]</td>
</tr>
<tr>
<td>Assignment Performance (e.g., grade, time to completion)</td>
<td>[52, 54, 59, 60, 65, 82, 92, 105, 106, 123, 124, 127, 143, 151, 165, 179, 187, 188, 195, 209, 216, 223, 228, 239, 242, 251, 256, 264, 286, 309, 310, 352, 364, 365, 376, 383, 389, 395, 399, 408, 415]</td>
</tr>
<tr>
<td>GPA or GPA Range (including CGPA, SGPA)</td>
<td>[4, 12, 28–32, 36, 37, 64, 66, 73, 113, 129, 135, 137, 140, 149, 167, 181, 189, 206, 211, 218, 240, 249, 255, 263, 265, 266, 268, 291, 297, 322, 325, 339, 344, 345, 366, 383, 390, 402, 405, 413]</td>
</tr>
<tr>
<td>Course Grade Range (e.g., A-B/C-F, Pass/Fail)</td>
<td>[17, 20, 38, 52, 58, 61, 77, 95, 108, 110, 111, 121, 126, 130, 139, 172, 174, 182, 194, 207, 222, 226, 227, 238, 254, 257, 258, 260, 267, 278, 289, 293, 316, 317, 323, 331, 332, 361, 363, 370, 379–381, 386, 402, 403, 406, 418, 421]</td>
</tr>
<tr>
<td>Knowledge Gain</td>
<td>[23, 156, 164, 192, 201, 270, 401, 419]</td>
</tr>
</tbody>
</table>

Table 14: Checklist for a minimum viable article on predicting student performance

<table>
<thead>
<tr>
<th>Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Define what is being predicted. If the value describing performance (e.g., course grade) consists of multiple items (e.g., course exam, course assignments), describe the contribution (weight) of each item when the performance value is calculated.</td>
</tr>
<tr>
<td>Define the factors used for prediction. Describe them in such detail that a reader that is not familiar with your particular context understands them. If factors are intertwined (e.g., course assignments) with the predicted value (e.g., course exam), be explicit about the connection. Provide links to, or if not possible, include the scales and surveys that have been used when collecting data.</td>
</tr>
<tr>
<td>Define the methods used for prediction and link the methods used by referencing appropriate articles. Unless you propose a novel method, formal proofs, etc. are not required. If you use feature selection, include details on them.</td>
</tr>
<tr>
<td>Define the data. Explain where the data comes from, if it is self-reported or automatically collected and if students are compensated for participating. Moreover, if the data contains students from a course, discuss the number of students in the course, describe how many were excluded from the analysis and why, and provide descriptive statistics that outlines the data. Be specific on whether the data is from a single course, or single institution, and also discuss if a separate data set is used for validating the prediction results.</td>
</tr>
<tr>
<td>Provide the results. Perform and report on the tests necessary to test required attributes of the data. Name the analyses being performed and report all the relevant statistics to allow for interpretation of the results.</td>
</tr>
<tr>
<td>Discuss the reasons why specific factors, performance metrics and methods were chosen (or omitted).</td>
</tr>
<tr>
<td>Reflect upon the results and consider why the methods and factors used did work or did not work. What are the particular context-specific issues that may influence the outcomes?</td>
</tr>
<tr>
<td>Describe threats to validity and limitations. Note situations in which a model or approach might be applied as well as where it is not valid.</td>
</tr>
</tbody>
</table>

- Initial Vocabulary/Taxonomy for Population
  - K-12 (from kindergarten to high-school)
  - Minors
  - Majors
  - Non-majors
  - Professional
  - Unknown / Unclear / Vague
  - Graduate (e.g. MSc students)
  - Undergraduate (e.g. BSc students)
  - Not applicable
  - Other

- Initial Vocabulary/Taxonomy for Training mode
  - Blended
  - Local
  - MOOC

- Online
- Unclear / vague
- Not applicable
- Other

- Initial Vocabulary/Taxonomy for Education Type
  - Formal education (e.g. university education, high-school)
  - Informal education (e.g. MOOC, work-related training, ...)
  - Not applicable

- Initial Vocabulary/Taxonomy for Data set information
  - Data coming from a single institution
  - Data coming from multiple institutions
  - Data coming from a single course
  - Data coming from multiple courses
  - Data is a publicly available dataset
  - Other
• Initial Vocabulary/Taxonomy for Family background
  - Family background in General
  - Income
  - Status
  - Support
  - Parent Educational Level
  - Parent Occupation
  - Number of Siblings
  - Caretaker Role
  - Not applicable
  - Other

• Initial Vocabulary/Taxonomy for Demographic data
  - Demographic data in General
  - Accessibility (disability)
  - Gender
  - Age
  - Ethnicity
  - International
  - Minority
  - Not applicable
  - Other

• Initial Vocabulary/Taxonomy for Working Conditions
  - Working Conditions in General
  - Distance to school
  - Daily commute time
  - Access to internet
  - Basic needs (water and toilet)
  - Not applicable
  - Other

• Initial Vocabulary/Taxonomy for Education background
  - Education background in General
  - Accommodations (modified programs, disability)
  - Extra-curricular activities
  - Background in CS
  - Background in Math
  - Background in Physics
  - GPA (high school)
  - GPA (post-secondary)
  - GPA (unknown source)
  - High School Quality
  - Previous (hobby) experience on topic
  - Previous (work) experience on topic
  - Standardized graduate admission test (e.g., GRE)
  - Standardized undergraduate admissions test (e.g., SAT, ACT)
  - Background in other disciplines
  - Grades from previous courses (at the post-secondary institution)
  - Grades from previous courses (before post-secondary institution)
  - Education background Not applicable
  - Education background: Other

• Initial Vocabulary/Taxonomy for Course (current or parallel)
  - Course in General
  - Attendance
  - Activity (with learning resources or materials)
  - Activity (with discussion forums or chats)
  - End of term assessment (exams)
  - Time on task
  - Marks
  - Marks (tests and quizzes)
  - Marks (assignments)
  - Marks (lab work)
  - Midterm assessment (tests, quizzes, midterm exams)
  - Pre-test score / mark

  - Organizational: Pedagogical methods
  - Organizational: Teaching mode
  - Organizational: Materials Available (what types?)
  - Social: Related social data (e.g., number of friends in class)
  - Social: Unrelated social data (e.g., Facebook, Twitter)
  - Not applicable
  - Other

• Initial Vocabulary/Taxonomy for Student motivation
  - Motivation in General
  - Desired grades
  - Extrinsic
  - Intrinsic / interest / passion
  - Importance of Grades
  - Utility
  - Not applicable
  - Other

• Initial Vocabulary/Taxonomy for Psychological/affective/learning
  - Scales in General
  - Achievement goals
  - Emotional (self-worth, anxiety, etc)
  - Goal-orientation (performance, mastery, etc)
  - Grit
  - Self-efficacy
  - Learning Strategies (deep, surface)
  - Learning Styles (!)
  - MSLQ
  - Personality
  - Self-regulation
  - Not applicable
  - Other

• Initial Vocabulary/Taxonomy for Research Method type
  - Mixed-methods
  - Qualitative
  - Quantitative
  - Not applicable
  - Other

• Initial Vocabulary/Taxonomy for classification (supervised learning)
  - Classification in General
  - Neural network
  - Adaptive boosting
  - SMOTE
  - Radial Basis
  - Naive Bayes
  - Nearest Neighbor
  - Decision Trees
  - Random Forest
  - SVM
  - Knowledge modeling
  - Bayesian network
  - Other

• Initial Vocabulary/Taxonomy for Clustering (unsupervised learning)
  - Clustering in General
  - Neural network
  - K-means
  - K-star
  - Other

• Initial Vocabulary/Taxonomy for Mining (finding frequent patterns/feature extraction)
  - Mining in General
  - Distributed
  - Text
  - Web
Predicting Academic Performance


[102] José R Deluz, Rolando Garcia Juan C. Morales, and Griselda Pérez. 2015. Markov chain modeling: student flow through remedial Mathematics to sophomore ISERC. In IEEE JUSE Annual Conference. Institute of Industrial and Systems Engineers (IISE), 3171.


Predicting Academic Performance

In ASEE Annual Conference & Exposition. American Society for Engineering Education.


Tyn Huang, Ian Do Guerrra-Hollstein, and Peter Brusilovsky. 2016. Modeling Student Combination Patterns for Deeper Knowledge Tracing. In UMAP Extended Proceedings.


In ASEE Annual Conference & Exposition. American Society for Engineering Education.


Qiang Hao, Ewan Wright, Brad Barnes, and Robert Maribe Branch. 2016. What are the most important predictors of computer science students’ online help-seeking behaviors? Computers in Human Behavior 62 (2016), 467–474.


Priyanka Anandrao Fatil and RV Mane. 2014. Prediction of Students Performance Using Frequent Pattern Tree. In 2014 International Conference on Computational Intelligence and Communication Networks (CICN) IEEE, 1078–1082.


Predicting Academic Performance


Sam C. Webb. 1951. A generalized scale for measuring interest in science subjects. Social Networks and Analysis 11, 3 (1951), 456–469.


Edward Tovar and Oliver Soto. 2010. The use of competences assessment to determine the impact of demo-


