Article

Modeling e-Behaviour, Personality and Academic Performance with Machine Learning

Serepu Bill-William Seota[1], Richard Klein[2] and Terence van Zyl[3]

[1] University of the Witwatersrand; bseota@gmail.com
[2] University of the Witwatersrand; richard.klein@wits.ac.za
[3] University of Johannesburg; tvanzyl@uj.ac.za

Featured Application: The e-Behaviour, personality and performance evaluation frameworks described in this article can be used by students and academic staff alike to monitor performance and online behaviour as it relates to performance. Being aware of e-Behavioural patterns is a starting point to improving the academic performance of individual students and groups of students. The methodology can be used to inform the extent to which a course is to be adapted such that it encourages students to engage in behaviour that promotes better academic performance.

Abstract: The analysis of student performance involves data modelling that enables the formulation of hypotheses and insights about student behaviour and personality. We extract online behaviours as proxies to Extraversion and Conscientiousness, which have been proven to correlate with academic performance. The proxies of personalities we obtain yield significant ($p < 0.05$) population correlation coefficients for traits against grade – 0.846 for Extraversion and 0.319 for Conscientiousness. Furthermore, we demonstrate that a student's e-behaviour and personality can be used with deep learning (LSTM) to predict and forecast whether a student is at risk of failing the year. Machine learning procedures followed in this report provide a methodology to timeously identify students who are likely to become at risk of poor academic performance. Using engineered online behaviour and personality features, we obtained a classification accuracy (κ) of students at risk of 0.51. Lastly, we show that we can design an intervention process using machine learning that supplements existing performance analysis and intervention methods. The methodology presented in this article provides metrics that measure the factors that affect student performance and complement existing performance evaluation and intervention systems in education.

Keywords: e-Behaviour, Big-five personality, Student performance

Citation: Seota, S.; Klein, R.; van Zyl, T. Modelling e-Behaviour, Personality and Academic Performance with Machine Learning. *Appl. Sci.* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

The evaluation and analysis of the factors that affect the performance and result of students stem from a need to improve student throughputs. Richiţeanu-Năstase and Stăiculescu [1] identify several reasons why post-secondary educational institutions have a low rate of completion. They name three main reasons: first, a lack of support, second, the student's background, and third, an inability to adapt to the curriculum.

In addressing student performance, we consider their grades at the end of a study programme as a measure of their performance. We also refer to performance as *risk* or *risk of failure* since an increase in performance results in a lowered risk of failure. The *e-Behaviour* of a student is "a pattern of engagement with a Learning Management System (LMS)", and *personality* Wright and Taylor [2] refers to "[...] the relatively stable and enduring aspects of individuals which distinguish them from other people and form the basis of our predictions concerning their future behaviour."

Traditional approaches to revealing relationships between student behaviour, personality and performance include questionnaires, surveys and interviews. Respondents' biases from the above qualitative methods of data collection can compromise the accuracy of their responses [3–5]. Furthermore, it has not proven easy to measure the reliability of an opinion [6], especially for each individual in a population. We address the self-reporting problems using unobtrusive and automated approaches that measure how students behave rather than how they think they behave. For instance, instead of asking, 'In how many weekly online discussions do you participate?', we instead obtain the exact number of discussions from an LMS register. The models developed in this research use quantitative metrics to proxy behaviour and personality traits traditionally obtainable from surveys. These metrics are used to draw correlations with features later used to predict student performance. From e-Behaviour and personality, we modelled an intervention framework that supplements current student intervention systems.

We extract behavioural insights linked to two of the five personality traits in the Big Five personality model through quantitative analysis – Conscientiousness and Extraversion. We use statistical metrics to extract forum and login behaviours, respectively. We define the relationships between these metrics and online behaviours, detail the relationships between a student's expressions of personality traits through their behaviours. Through this research, we:

1. define a framework for personality traits and behaviour in the context of student online engagement,
2. show the relationship between Bourdieu's Three Forms of Capital and academic performance,
3. show the relationship between personalities and academic performance through e-behaviours,
4. show that we can use e-Behaviour and machine learning to predict student performance, and
5. highlight the importance of the explainability of modelled personality traits and e-Behaviours.

This work contributes to the prediction of student outcomes using online behaviours in the following ways:

- We present a framework and methodology for arriving at predictive models for student performance that starts with personality traits. These traits are the drivers of online behaviours that generate features that are predictive of performance.
- We argue for the use of online behaviours and proxies for the personality traits: Conscientiousness and Extraversion.
- We demonstrate that online behaviours that are strongly associated with the identified personality traits correlate with student performance in a statistically significant way.

1.1. Literature Review

Factors that affect a student's performance do not only relate to their cognitive ability. Psychological traits also play a role in determining student success [7]. As a result of the intricacies underlying these traits, common approaches to demonstrating relationships between psychological traits and academic performance often involve surveys given to each student, as was done in studies by Poropat [7], Morris and Fritz [8], Kim *et al.* [9]. A qualitative survey, as efficiently designed as it may be, is prone to biases of its participants Heppner *et al.* [3], Stone [4], Northrup *et al.* [5].

One standardised psychological framework for measuring personality is the Five-Factor or OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism) model Poropat [7], Morris and Fritz [8], Costa and McCrae [10], Furnham *et al.* [11], Ciorbea and Pasarica [12], Kumari [13]. Recent research by Morris and Fritz [8] has shown that of these five factors, Conscientiousness and Extraversion are highly correlated with student educational outcomes. In this research, we build upon the vast body of literature that supports these two personality traits as being strongly correlated with student performance [7–9,11–14].

Revised Neuroticism-Extraversion-Openness Personality Inventory (NEO PI-R) [15] expands each of the OCEAN personality trait's six facets. For Conscientiousness, these facets are competence, orderliness, dutifulness, achievement-striving, self-discipline, and deliberation. Activity, assertiveness, excitement-seeking, gregariousness, positive emotion and warmth are the facets of Extraversion. Wilt and Revelle [16] define Extraversion as the 'disposition to engage in social behaviour', which links significantly to the gregariousness facet. *Gregariousness* is defined as the 'tendency for human beings to enjoy the company of others and to want to associate with them in social activities' [17]. *Dutifulness* is defined as the characteristic of being motivated by a sense of duty [18]. In this article, we model for the orderliness facet of Conscientiousness and the gregariousness facet of Extraversion.

Recent work by Akçapınar [19] and Huang *et al.* [20] has shown that the usage of online behaviours alone does not necessarily lead to features that are predictive of student performance. We argue that this may be due to the many features that can be engineered from the time series of log data representing online student behaviour. For instance, the Tsfresh library can extract over 40 time-series features. The link between online behaviours and the personalities that underlie them has not been extensively explored. We follow the argument of Khan *et al.* [21] and postulate that starting from a principled approach that is grounded in personality traits will lead to a more viable set of features and metrics.

1.1.1. Contribution to Existing Evaluation Systems

The University of the Witwatersrand's (the University's) academic and student support staff members have access to three standard systems of identifying the likelihood of students completing a programme. These systems can be broadly grouped into grades, questionnaires, observing a student's grades for that programme over time, and one-on-one consultations by a counsellor or lecturer with the student.

Questionnaires have two significant limitations. Firstly, they are not offered throughout the teaching period and secondly, they are anonymous, meaning there is no easy way of linking students at risk to their programmes. Fowler and Glorfeld [22], Poh and Smythe [23], Evans and Simkin [24] show that prior performance or grades are a reliable measure for future performance. By their high-touch nature, observing grades and consultations are usually not anonymous. The advantage of these two systems is that they give a detailed response to students' feelings towards their programmes and are thus potentially corrective (can help resolve poor performance). The disadvantages are that one-to-one consultations and grades are often retroactive rather than proactive and not conducted at scale or sufficiently continuously.

The limitations in gauging student performance through the above mechanisms give rise to a proposal for using e-Behaviour machine learning models. While models that fit students' e-behaviour do not guarantee similar reliability, e-Behaviour machine learning models have some advantages over grades, questionnaires and consultations. Table 1 suggests that e-Behaviour models are the only system of evaluation that is continuously proactive – can be monitored at any point in time to take corrective action.

Table 1: Comparison of Evaluation Systems

System	Continuously Proactive	Corrective	Easily Feasible at Scale	Reliable
Questionnaires	✗	✗	✓	✗
Previous Grades	✗	✗	✓	✓
Consultation	✗	✓	✗	✓
e-Behaviour Models	✓	✗	✓	To be shown

1.2. Bourdieu's Three Forms of Capital and Student Success

Bourdieu's Three Forms of Capital is a framework that suggests that economic, cultural and social capital that an individual can leverage regulates her level of success. We use this framework to support our investigation of the economic, cultural and social capital that a student has available to her as each form of capital relates to her academic performance. Dauter [25] defines economic sociology as

> "[...] the application of sociological concepts and methods to the analysis of the production, distribution, exchange, and consumption of goods and services."

Economic sociology has been used extensively by Bourdieu and Richardson [26], who argue that an individual's possession of three forms of capital regulates her social positions and ability to access goods and services. These three forms of capital are economic capital, cultural capital, and social capital. The Three Forms can be considered essential to a student obtaining good grades and acquiring the services she needs to improve her grades. We refer to our proxies for economic and cultural capital as the *background* of a student.

1.2.1. Social Capital

Bourdieu and Richardson [26] define social capital as:

> "The aggregate of the actual or potential resources which are linked to the possession of a durable network of more or less institutionalised relationships of mutual acquaintance and recognition."

The above definition describes social capital as a resource that is available between people due to their relationships. An individual may accrue social capital by being part of relationships. Carpiano [27] uses the framework by Bourdieu and Richardson [26] to build onto the theory of social capital. Carpiano [27] categorises the social capital available to individuals into four types, namely: social support, social leverage, informal social control, and community organisation participation.

The above four types of social capital are available to students, forming relationships for social or academic purposes. Hallinan and Smith [28] refer to these intra-cohort groups as *social networks* or *cliques*. The common saying, *show me your friends and I will show you your future*, is commonly used to describe the relationship between an individual's affiliates and her results. In this research, these results are referred to as her *Grade* or her *Outcome*. The hypothesis that a student has access to some *social capital* has been validated to various extents by Hallinan and Smith [28] and is also adopted in this research.

A limitation with the social capital frameworks by Bourdieu and Richardson [26], Carpiano [27] and Song [29] is that they provide no standard measure of social capital. The definition of social capital leaves no room for a well-defined metric. In our research, a student's social network is evidence of her social capital and is called her *Academic-group*. In an academic setting, a student's *quality* of resources social capital can be defined in terms of the aggregate grades of her Academic-group. The relationships between Academic-groups and Grades is modelled in Sections 3.4 and 3.5.

1.2.2. Cultural Capital

According to Hayes [30], cultural capital is a set of non-economic factors that influence academic success, such as family background, social class and commitments to education, and do not include social capital. Bourdieu and Richardson [26] categorises cultural capital into three forms, namely:

1. institutionalised cultural capital (highest degree of education),
2. embodied cultural capital (values, skills, knowledge, tastes), and
3. objectified cultural capital (possession of cultural goods).

In this research, the features we selected in Section 2.5 are proxies of 1 and 2. Smith and White [31] found that success in obtaining a degree relates strongly to gender and ethnicity. Caldas and Bankston [32] found that students' cultural capital affect their performance.

1.2.3. Economic Capital

Bourdieu and Richardson [26] defines *economic capital* as material assets that are 'immediately and directly convertible into money. In turn, an individual's monetary leverage can be converted into cultural and social capital [26].

Bourdieu and Richardson [26] recognise that an individual can increase her social and cultural capital by making use of her economic capital. An individual who leverages her economic capital can obtain more resources to improve her cultural capital. For instance, an individual can improve her cultural capital through improvement in her position in society. By investing in formal or informal education beyond the classroom, a student may increase her knowledge and the amount of cultural capital available to her. Fan [33] observed that a student's quality and level of education was affected by her cultural and economic capital.

Section 3.1 reveals the relationships between student background (*background* refers to cultural and economic capital) and academic performance.

2. Methodology

2.1. Data Preprocessing

The data is made up of files with logs on the *Moodle* LMS database at our university for first-, second-and third-year students who were enrolled in Applied Mathematics and Computer Science modules in the 2018 academic year. After examining distributions and removing students who had no grade records, the data was reduced to time series patterns, aggregated by each day of the semester. The models were fitted on the open-semesters logs recorded (from the beginning of the semester till two weeks before exams). The advantage of using only open-semester data is not only that it helps us understand the predictive power of the behaviour, but it also eliminates effects of sudden changes in behaviours that are forced upon students as examinations approach [8]. The target variable for all experiments is the aggregate Grade (out of 100 points) of online assessments, including examinations, that the student obtained over the year.

2.2. Importance and Choice of Personality Traits

The university LMS contains several tables that each provides different information. We checked each table's appropriateness in modelling any of the OCEAN traits. Extraversion and Conscientiousness were traits linked closest to the information in the tables and were chosen for this study. By comparison, Openness, Agreeableness and Neuroticism were more complex to model, given the available data and the lack of validation of a link to academic performance within the literature.

Our data linked closest to the dutifulness facet of Conscientiousness and the gregariousness facet of Extraversion. Alternative formulations of each trait were considered and are described in Section 5.2. We used quantitative proxies to model Conscientiousness (Dutifulness) and Extraversion (Gregariousness).

2.3. Encoding Personality Traits

According to Ajzen [34], Campbell [35], human behaviour can be explained by reference to stable underlying dispositions or personality. Wright and Taylor [2] define personality as:

> 'the relatively stable and enduring aspects of individuals which distinguish them from other people and form the basis of our predictions concerning their future behaviour'.

Therefore, Extraversion and Conscientiousness were modelled as single-valued averages that do not vary through time. Our choice to encode personality traits as unvarying values are based on theory by Wright and Taylor [2], Ajzen [34], Campbell [35], Hemakumara and Ruslan [36].

2.3.1. Challenges against encoding personality traits

The above definitions of personality and their link to behaviour may cause a belief that personality and behaviour should be measured identically. However, to understand the separate correlations between either personality and performance or e-Behaviour and performance, it is essential to encode an individual's *stable aspects* (personality traits that are less likely to change) differently from her *changing* e-Behaviour. As a result, personality metrics are aggregated while e-Behaviour is modelled to vary over time.

2.4. Encoding Performance

Three measures of performance were constructed, namely *Grade*, *Outcome* and her derived *Safety Score*. Grade is a continuous label that indicates the mean of a student's performance across all modules taken. This label is continuous and ranges between 0.00 and 100.00. Outcome is a binary label that indicates whether a student obtained below 51 *Grade* points (*At-risk*) or at least 51 *Grade* points (*Safe*). That is, the Outcome is taken to measure the degree of *risk-of-failure*. Note that a *fail* is considered any grade below 50 Grade points. However, the boundary of 51 provides a buffer that allows the models to reveal students who were close to failing (At-risk). Therefore, a student need not fail for her to be considered at risk. A student's Safety Score is a classification label used as a label of her predicted Outcome. A correct classification would assign a *Flagged* Safety Score for an At-risk student and an *Ignored* Safety Score for a student with a Safe Outcome.

Grade is used as a regressor against Extraversion-level (Section 2.6) and Conscientiousness-level (Section 2.9). *Outcome* is used as a label to the classification models in Sections 2.5 and 2.10.

2.5. Student Background

The raw Background Dataset consists of 4 748 students and 176 features on which experiments were conducted. These features captured answers by the student upon registration and data collected throughout their study – for instance, their high-school facilities, high-school subjects, age and city of residence. Table 2 shows a summary of the features after each phase of transformation.

Table 2: Background Data Feature Count Per Phase of Transformation

Transformation Phase	**Categorical Features**	**Total Features**
Before One-hot	169	176
After One-hot	6 616	6 623
After RFE	5	5

The 169 categorical features were one-hot encoded, extending the number of features from 176 to 6 623. Recursive Feature Elimination algorithm (RFE) with a Decision Tree was used to reduce the 6 623 feature set's dimensionality.

2.5.1. Feature Selection using RFE

RFE involves filtering through features with the lowest ranking of importance against Outcome, through the following procedure [37]:

1. optimise the Decision Tree weights with respect to its objective function on a set of features, F
2. compute the ranking of importance for the features in F using the Decision Tree optimiser
3. prune the features with the lowest rankings from F
4. repeat 1 - 3 on the pruned set until the specified number of features is reached.

The RFE process produced five Background Features, explained with the Grade and Outcome variables in Table 3.

Table 3: Background Data Features and Labels after RFE

Feature	**Description**	**Type**	**Values**
Quintile	To which of the five categories a student's high-school belongs under the South African Government schools standards; a 6 indicates private high-schools	Categorical	1 - 6
Gauteng Province	Whether a student completed their ultimate year of high school at a school in *GP* (Gauteng Province)	Binary	No, Yes
Gender	Whether the student is female or male	Binary	Female, Male
Financial Assistance	Whether a student received financial aid from the National Student Financial Aid Scheme	Binary	No, Yes
Township School	Whether a student's high school was situated in a township area	Binary	No, Yes
Grade (label)	Grade points out of 100 obtained, as defined in Section 2.4 on Encoding performance	Continuous	0.0-100.0
Outcome (label)	Risk of the student based on their Grade, as defined in Section 2.4 on Encoding performance	Binary	At-Risk, Safe

2.6. Extraversion and Academic-groups

Discussion, *Message* and *Time* independent variables, explained in Table 4, were used to engineer the Extraversion-level of a student, as well as formulate the Discussions and Collaboration-groups.

Table 4: Forum Table Features

Feature	**Description**	**Type**	**Possible Values**
Discussion	Discussion Number. Messages that begin a topic and are posted as responses are assigned the same discussion number.	Categorical	0 – 337
Message	Contents of each forum post.	String	-
Time	Extracted from the *Created* variable. Indicates the time at which each *message* was posted.	yyyy-mm-dd	2018-01-05 to 2019-01-05
Grade (label)	Number of points out of 100 (Section 2.4).	Number	0.0 – 100.0

2.6.1. Forum Posts and Extraversion:

The definition of social capital in Section 1.2 suggests that Extraversion or gregariousness can improve an individual's ability to accumulate social capital, which is correlated with academic performance. A way to model social interaction or gregariousness is by capturing the number of forum posts that an individual contributes to forum discussions. Hence, we chose the student's post count as a quantitative proxy for her *level* of Extraversion.

Each student was placed in an Extraversion-level group, E, representing the number of posts she contributed. Each level, E, was then assigned a mean Grade, G_E, computed by averaging the grades of all students in E.

Table 5: Input Table – Extraversion-level Grade against Extraversion-level

E	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
$\mathbf{G_E}$	55.8	64.4	68.3	66.1	69.3	73.3	66.0	69.7	71.4	70.2	81.3	79.4	75.2	74.4	79.3

2.7. Student Discussions

A Discussion (group), d_i, is defined as any discussion that contains more than two students created on the Moodle LMS. A Discussion that has fewer than three students is not considered a Discussion by our definition. Linear OLS assumptions for Discussions containing only two or more students do not hold. Section 4.4.1 shows the reasons. Let $D = \{d_i\}_{i=1}^{k} = \{d_1, d_2, \ldots, d_k\}$ be the set of all Discussions, s_j represent any student who participates in discussion d_i, s_i represent a selected random student who participates in discussion d_i, $\mathbb{E}[Gd_i]$ represent the mean Grade of Discussion d_i, and $G(s_j)$ is the grade of student s_j.

$$\mathbb{E}[Gd_i] = \frac{1}{n(d_i) - 1} \sum_{s_j \neq s_i} G(s_j), \tag{1}$$

where k represents the number of Discussions in D and $n(d_i)$ denotes the number of students in d_i. This section measures the correlation between $Gd_i(s_i)$ and $\mathbb{E}[Gd_i]$ by following Algorithm 1:

Algorithm 1: Correlation Between Mean Discussion Grade and Student Grade

Result: $\hat{G}d_i(s_i) = \beta_0 \mathbb{E}[Gd_i] + \beta_1$

foreach $d_i \in D$ **do**
  Select s_i
  Obtain $Gd_i(s_i)$ from s_i
  Compute $\mathbb{E}[Gd_i]$
  Plot $(Gd_i(s_i), \mathbb{E}[Gd_i])$
end

2.8. Student Collaboration-groups

This section illustrates an alternative method to formulating an Academic Group, namely, the Collaboration-group method. We correlate the Grades of students within each Collaboration-group with the mean Grade of each Collaboration-group.

The raw Forum Table was transformed into Table 6 below, which shows discussion participation per student. Each column, d_i, represents a discussion. 1 represents that the student participated in discussion d_i, while 0 shows that she did not participate in d_i.

Table 6: Sample Table of Discussion Participation

s	$\mathbf{d_0}$	$\mathbf{d_1}$	$\mathbf{d_2}$	$\mathbf{d_3}$	$\mathbf{d_4}$	...	$\mathbf{d_{337}}$
1	0	0	0	0	0	...	1
2	1	1	0	0	0	...	0
3	0	0	1	1	0	...	0
...	...	...	...	...	...	...	...
1131	0	1	0	1	0	...	0
1132	1	0	0	0	0	...	0
1133	0	1	0	0	0	...	1

Let $C = \{c_i\}_{i=1}^{k} = \{c_1, c_2, \ldots, c_k\}$ be the set of all Collaboration-groups. h_i is the *Host* of c_i with $H = \{h_i\}_{i=1}^{k} = \{h_1, h_2, \ldots, h_k\}$ be the set of all Hosts, one for each Collaboration-group.

A Collaboration-group, c_i, that is *hosted by* student, h_i, is defined as the group of more than two students with whom h_i shares at least one discussion. Any group with two or fewer students is not considered a Collaboration-group by our definition; OLS relationships analogous to those in this section do not hold for groups containing only two or more students. The reasons are presented under Section 4.4.1. h_i may host a maximum of one Collaboration-group. Let $\mathbb{E}[Gc_i]$ represent a the mean Grade of c_i, and $\mathbb{E}[Gc_i]_i(h_i)$ represent the Grade of h_i where $\mathbb{E}[Gc_i]$ represents the mean Grade of c_i which excludes $Gc_i(h_i)$, as in the case with $\mathbb{E}[Gd_i]$ and $Gd_i(s_i)$ in Equation 1.

The kNN algorithm was used to compute the Collaboration-group for each student, using the **Collaboration-group policy** specified in the below paragraph. By this policy, not all students fit the qualify to host a Collaboration-group.

We design the conditions necessary to define the Collaboration-group policy; let h_* be a *candidate* Host of a Collaboration-group, with c_* representing the Collaboration-group to be hosted by student, h_*. $n(c_*)$ represents the number of students in c_*, s_1, s_2 and s_3 are any three students in the cohort, and h_i represents a (qualified) Host to her (unique) Collaboration-group, c_i.

Collaboration-group Policy: c_* becomes a Collaboration-group, c_i, if and only if $n(c_*) >$ 2 students. Equivalently, if h_* shares a discussion with s_1, s_2 and s_3, then h_* *qualifies* as a Host, h_i, and $c_i = \{s_1, s_2, s_3\}$. If $n(c_*) \leq 2$ students, then h_* remains a candidate until she shares a discussion with at least one more member.

A sample set of the Hosts, $\mathbf{h_i}$, and their Collaboration-groups, $\mathbf{c_i}$ is shown in Table 14. Each entry in column $\mathbf{c_i}$ is a set of indices that represent students in c_i, while column $\mathbf{Gc_i(h_i)}$ shows the Grades of the Hosts. $\mathbb{E}[\mathbf{Gc_i}]$ represents the mean Grades of each c_i.

2.9. Logins and Conscientiousness

Section 2.2 explained the facets that describe each personality trait. Our model of Conscientiousness relates closely with dutifulness. Barrick *et al.* [38] and Campbell [39] theorise that Conscientiousness is linked to an individual's choice to expend a level of effort. Therefore, modelling dutifulness requires a formulation that captures the average logins per week that the student made throughout the programme. This model of Conscientiousness-level captures the facet of dutifulness and the choice to expend effort. Let $C(s)$ be a variable that represents the Conscientiousness-level of student s. $C(s)$ is modelled as the average number of logins over the period spanning a student's active weeks. For each student, $C(s)$ is formulated as:

$$C(s) = \frac{\sum_t^{17} L(s)_t}{W(s)}, \tag{2}$$

where $W(s)$ is the number of weeks spanned between the student's first and last login.

The reason for modelling $C(s)$ as the *average* number of active days per week instead of the *total* number of logins over the period is that the *average* normalises the data. Averaging reduces biases caused by differences in the number of days per cohort, per subject and programme, that students are expected to log in.

Each personality trait proxy was regressed against the students' grades for the semester using the Ordinary Least Squares (OLS) regression method [40]. The associated slope and correlation coefficients, p-values and Slope coefficient with 95% confidence intervals are reported.

To date, several longitudinal studies investigating academic performance and personality use *effects, determines* or *predicts* to mean the *relationship* or *correlation* between personality traits and performance (for example, works by Poropat [7], Chamorro-Premuzic and Furnham [14], Blumberg and Pringle [41] and Morris and Fritz [8]. Without knowing the causality of personality traits on performance in our study, we adopt the same terminology for ease of reference and comparison.

2.10. Behaviour-Personality and Behaviour Model

The Behaviour-Personality model (B-PM) consists of two components: the behavioural component is the Login Sequences of students, while the personality component augments the Login Sequences. The traits that make up the personality component are the Extraversion- and Conscientiousness-levels. The Behaviour-Model (BM) consists of only the Login Sequences of students as input.

For each student s, we engineered the Login Sequence ($\{L(s)_t\}$), Extraversion-level ($E(s)$), and Conscientiousness-level ($C(s)$ by augmenting $\{E(s)_t\}_{t=1}^{17}$ and $\{C(s)\}_{t=1}^{17}$ as *sequences* of the *same* values that run parallel to $\{L(s)_t\}_{t=1}^{17}$ through time t. This augmentation forms a 3×17 *input array of sequences*:

$$[\{L(s)_t\}, \{E(s)_t\}, \{C(s)_t\}],$$

where $\{L(s)_t\}_{t=1}^{17}$ is a sequence of values that *vary* through time t, $\{E(s)_t\}_{t=1}^{17}$ is a sequence of the *same* value through time t, so that $E(s)_t = E(s)_{t-1}$ for all Whole Numbers $t \in [2, 17]$, and $\{C(s)_t\}_{t=1}^{17}$ is a sequence of the *same* value through time t, so that $C(s)_t = C(s)_{t-1}$ for all Whole Numbers $t \in [2, 17]$. (See Table 7)

This method of augmenting inputs in parallel was guided by its usage in Leontjeva and Kuzovkin [42]. As a result, The B-PM **input** for each student is the *array of sequences*:

$$[\{L(s)_t\}, \{E(s)_t\}, \{C(s)_t\}].$$

The B-PM **output** for each student is a Safety Score: *Flagged* for At-risk students, and *Ignored* for Safe students. These per student B-PM **input** and **output** structures are summarised in Table 7.

Table 7: B-PM Training Input and Output Summary

Feature	Shape	Type	Example Value
$\{L(s)_t\}$ Login Sequence	(1×17)	A Sequence of Whole Numbers	$[3, 7, \ldots, 0]$
$\{E(s)_t\}$ Extraversion[1] Sequence	(1×17)	A Sequence of Whole Numbers	$[8, 8, \ldots, 8]$
$\{C(s)_t\}$ Conscientiousness[2] Sequence	(1×17)	A Sequence of Real Numbers	$[1.1, 1.1, \ldots, 1.1]$
Input: $[\{L(s)_t\}, \{E(s)_t\}, \{C(s)_t\}]$	(3×17)	An *Array of Sequences* of Real Numbers	$[[3, 7, \ldots, 0]$ $[8, 8, \ldots, 8]$ $[1.1, 1.1, \ldots, 1.1]]$
Output: Safety Score = $\{Flagged, Ignored\}$	(1×1)	Binary	Ignored

2.11. Algorithms for e-Behaviour, Personality and Performance Analysis

2.11.1. Decision Tree Classifier

The Decision Tree Classifier (DTC) is a supervised learning algorithm that iteratively assesses conditions on the values of features in a data set to perform classification. DTC breaks down a decision-making process into a collection of simpler decisions, providing classifications that are easier to interpret than other statistical and machine learning models [43].

DTC Architecture

DTC is assembled from a root node, edges, internal nodes and leaf nodes. At the root node, DTC conducts a test on each observation's value. Based on its value, the root node assigns a resolution represented by an edge, which the observation then traverses. At the end of the traversed edge is an internal node. An example of a node's test is 'Gender?', and an example of an edge is 'Female'. This decision process continues through the rest of the internal nodes until the tree reaches a leaf node, where a classification is made. See Mitchell [44] for details on the DTC architecture.

Gini Impurity Index – Decision Factors

During prediction, an observation is predicted as part of a class after being checked through a series of conditions. An optimal decision tree results in an *optimal split*. An optimal split is achieved when each leaf node has the fewest possible train-set misclassifications (lowest impurity), and the tree has not been overfitted. Entropy and Gini Impurity Index are two commonly used metrics for impurity. The Gini Impurity Index (Gini) measures the relative frequency that a randomly chosen element from that set would be mislabelled. A Gini score greater than zero describes a node that contains samples belonging to different classes. Raileanu and Stoffel [45] suggest that the difference between Entropy and Gini is trivial. This research uses Gini, which is interpretable.

Gini Calculation:

The Gini value decreases as a traversal is made down the tree towards its leaf nodes. The decrease happens as each internal node's condition aims to separate the classes according to a criterion that results in more homogeneous separations and higher accuracy in the training data. However, as with other predictive models, a high training-set accuracy is generated at the risk of overfitting. A larger tree (with more edges and branches) is more likely to overfit than a smaller tree and may result in a Gini of 0 at the tree's leaf nodes. A Gini of 0 represents the minimum probability of misclassification over the training set but may result in weak generalisability over the test set. Therefore, smaller trees are preferred to larger trees [44].

The Gini impurity index is calculated using the formula:

$$Gini = 1 - \sum_i p_i^2 \tag{3}$$

where p_i is the probability of class i.

Khalaf *et al.* [46] model DTCs on survey questions and answers that cover health, social activity and relationships of students to predict their academic performance. Topîrceanu and Grosseck [47] and Kolo *et al.* [48] provide literature in educational data mining and advocate for the use of the DTC due to its low complexity (with a run time of $O(m \times n \times log(n))$ and high interpretability. In Section 3.1.1, the DTC is used to select student economic and social capital features and predict the student Outcomes.

2.11.2. Ordinary Least Squares Linear Regression Analysis

Ordinary Least Squares (OLS) Linear Regression is a statistical model that estimates the linear relationship between one or more independent variables (regressors) and a dependent variable (regressand) [49]. Throughout this research, only one independent variable was used per regression model. Using one independent variable per model isolates the effect of each variable on Grade. A Regression model with one independent variable is called a Simple OLS Regression model. Each estimated or predicted value, $\hat{y}_i$, derived from the line of best-fit shown in Equation 5, can be determined by

$$\hat{y}_i = \beta_0 x_i + \beta_1 + \epsilon_i, \tag{4}$$

where $\hat{y}_i$ is the predicted value of the i^{th} independent variable, x_i. β_0 is slope coefficient of the model, representing the average marginal change in $\hat{y}_i$ for a unit increase in x_i. β_1, the intercept of the model, represents the expected value of $\hat{y}_i$ when $x_i = 0$. $\epsilon_i \in \mathbb{R}$ is the residual term.

Every observed value, y_i, has an associated estimate or prediction value, $\hat{y}_i$. The line of best-fit,

$$\hat{y} = \beta_0 x + \beta_1, \tag{5}$$

is obtained by minimising the sum of the squares in the difference between the observed and predicted values of the dependent variable,

$$\sum (y_i - \hat{y}_i)^2 = (y_i - (\beta_0 x_i + \beta_1))^2. \tag{6}$$

The (linear) correlation coefficient between x and y is represented by r or $r(x,y)$. $r(x,y)$ measures the extent to which the independent variable, x is correlated with the dependent variable, y. That is, $r(x,y)$ measures the degree of closeness of all points, (x,y), to the line of best-fit, $\hat{y} = \beta_0 x + \beta_1$. The correlation coefficient lies between -1 and 1, where a r of 1 or -1 means that the change in y is directly proportional to the increase in x. In that case, x and y are said to be perfectly correlated. That is,

$$r = 1 \implies \frac{y_i - y_{i+1}}{(x_i + 1) - x_i} = c, \tag{7}$$

for all values of i where x_i and y_i are defined, and where $c \in \mathbb{R}$. r is computed by

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \tag{8}$$

where $\bar{x}$ represents the mean average of independent variable x and $\bar{y}$ represents the mean average of dependent variable, y.

The p associated with β_0 shows the probability of a hypothetical value, β_0^*, having an absolute value, $|\beta_0^*|$, that is at least as high as the observed $|\beta_0|$ by chance. The level of significance, α, is used as a threshold for a permissible p. In the domain relating to e-Behaviour, personality and performance, the common α used is 0.05, or a 5% level of significance. In our regression models,

β_0 is accompanied by a ($1 - \alpha$ = 95%) confidence interval, $\beta_0 \pm v$. Suppose $p < \alpha = 0.05$ for β_0. This means that from 100 experiments on similar sample distributions, fewer than 5 experiments would produce a $|\beta_0^*|$ value that lies outside of $\beta_0 \pm v$. Such a result means that the regressor and regressand have a statistically significant correlation different from zero [49]. Statistical insignificance may indicate that x on its own does not yield reliable estimates, $\hat{y}$.

Statistical significance is important in analysing a student cohort's behaviour since statistical significance confirms the existence of a statistical relationship. Empirical significance refers to the magnitude of β_0 [49] and is also a measure of the model's practical value. One can be more confident in practical decisions if the relationship is not generated by chance (if the relationship is statistically significant). This *chance* is measured by the p.

2.11.3. Validity of OLS Regression Models

The data given in a model must satisfy five OLS regression assumptions [49], namely:

1. Normality of model residuals. The residual for each point is given by $y_i - \hat{y}_i$. $s^2 + k^2$ is computed for the residuals, where s is the z-score returned by the test for skewness and k is the z-score returned by the test for kurtosis.
2. Residual Independence or lack of Autocorrelation in Residuals.
3. Linearity in Parameters.
4. Homoscedasticity of Residuals.
5. Zero Conditional Mean.
6. No Multicollinearity in Independent Variables.

A linear relationship that violates the OLS assumptions is not fit for an OLS model. Therefore, we constructed only OLS relationships that satisfy the assumptions. Mention is made about experiments where the OLS assumptions are violated. Linearity in Parameters, Homoscedasticity of Residuals and Zero Conditional Mean were verified for all Regression experiments whose results were analysed. The No Multicollinearity assumption was not verified since all OLS regression experiments are Simple. See Gujarati and Porter [49] for further details on the formulation of the OLS Regression model.

2.12. Long Short-Term Memory

The Long-Short Term Memory algorithm (LSTM) is a deep-learning architecture designed to model sequences for prediction [50]. The LSTM has been used in studies that range from predicting weather-induced background radiation fluctuation by Liu and Sullivan [50], to human motion classification and recognition by Wang *et al.* [51].

The backpropagation through time algorithm computes the error, $\mathbf{E}_t$, at every time step, t, and then computes the total error. The LSTM's parameters are updated to minimise the total error $\frac{\partial \mathbf{E}}{\partial \mathbf{W}}$ with respect to a weight parameter $\mathbf{W}$:

$$\frac{\partial \mathbf{E}}{\partial \mathbf{W}} = \sum_{t=1}^{T} \frac{\partial \mathbf{E}_t}{\partial \mathbf{W}}. \tag{9}$$

Letting $\mathbf{y_t}$ represent the output at time t, $\mathbf{h_t}$ represent the hidden state at time t and applying the chain rule to the Recurrent Neural Network model, the total error in equation 9 becomes:

$$\frac{\partial \mathbf{E}}{\partial \mathbf{W}} = \sum_{t=1}^{T} \frac{\partial \mathbf{E}}{\partial \mathbf{y}_t} \frac{\partial \mathbf{y}_t}{\partial \mathbf{h}_t} \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_k} \frac{\partial \mathbf{h}_k}{\partial \mathbf{W}}, \tag{10}$$

where $\frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_k}$ involves a product of Jacobian matrices:

$$\frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_k} = \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_{t-1}} \frac{\partial \mathbf{h}_{t-1}}{\partial \mathbf{h}_{t-2}} \cdots \frac{\partial \mathbf{h}_{k+1}}{\partial \mathbf{h}_k}. \tag{11}$$

Equation 11 illustrates the problem of vanishing gradients in equation 9; when the gradient becomes progressively smaller as k increases, the parameter updates become insignificant.

LSTMs are an architecture of Recurrent Neural Networks (RNNs). Bengio *et al.* [52] suggest that RNNs are challenging to train because of the vanishing error gradient problem. The following section stipulates how the LSTM's architecture mitigates the vanishing error gradient issue through LSTM cells that maintain a state $\mathbf{c}_t$ at every iteration t. The cell state $\mathbf{c}_t$ serves to *remember* and propagate cell outputs between time steps. Each cell state then allows for temporal information to become available in the next time step, adding greater context to the inputs $\mathbf{x}_t$ that follow.

The activation $\mathbf{h}_t$ of an LSTM unit is:

$$\mathbf{h}_t = \mathbf{o}_t \tanh(\mathbf{c}_t), \tag{12}$$

where

$$\mathbf{o}_t = \sigma\big(\mathbf{W}_{xo}\mathbf{X}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o\big), \tag{13}$$

is an output gate that mitigates the amount of content in the memory to expose to the following time step and $\sigma : \mathbb{R} \to (0, 1)$ is the logistic sigmoid function.

Given new memory content,

$$\mathbf{i}_t \tanh\big(\mathbf{W}_{xc}\mathbf{X}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c\big), \tag{14}$$

where $\mathbf{i}_t$ represents the degree to which new memory is added to the memory cell, and is specified by an input gate

$$\mathbf{i}_t = \sigma\big(\mathbf{W}_{xi}\mathbf{X}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i\big), \tag{15}$$

the cell state,

$$\mathbf{c}_t = \mathbf{f}_t\mathbf{c}_{t-1} + \mathbf{i}_t \tanh\big(\mathbf{W}_{xc}\mathbf{X}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c\big), \tag{16}$$

can be updated by taking into account the previous cell state $\mathbf{c}_{t-1}$ and a term defined by the forget gate,

$$\mathbf{f}_t = \sigma\big(\mathbf{W}_{xf}\mathbf{X}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f\big). \tag{17}$$

Consolidating equations 12 to 17, the system of equations that describe each LSTM unit given by

$$\begin{align}
\mathbf{f}_t &= \sigma\big(\mathbf{W}_{xf}\mathbf{X}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f\big), \tag{18}\\
\mathbf{i}_t &= \sigma\big(\mathbf{W}_{xi}\mathbf{X}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i\big), \tag{19}\\
\mathbf{c}_t &= \mathbf{f}_t\mathbf{c}_{t-1} + \mathbf{i}_t \tanh\big(\mathbf{W}_{xc}\mathbf{X}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c\big), \tag{20}\\
\mathbf{o}_t &= \sigma\big(\mathbf{W}_{xo}\mathbf{X}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o\big), \tag{21}\\
&\text{and} \tag{22}\\
\mathbf{h}_t &= \mathbf{o}_t \tanh(\mathbf{c}_t). \tag{23}
\end{align}$$

Let B denote the input batch size (number of time stamps per input chunk), H denote the LSTM hidden state capacity, and D represent the dimensions of the inputs to the LSTM. Then, in equations 18 through 23:

$$\mathbf{x}_t, \mathbf{h}_{t-1} \in \mathbb{R}^{BxD}, \tag{24}$$
$$\mathbf{f}_t, \mathbf{i}_t, \mathbf{c}_t, \mathbf{o}_t, \mathbf{h}_t \in \mathbb{R}^{BxH}, \tag{25}$$
$$\mathbf{W}_{xf}, \mathbf{W}_{xi}, \mathbf{W}_{xc}, \mathbf{W}_{xo} \in \mathbb{R}^{DxH}, \tag{26}$$
$$\mathbf{W}_{hf}, \mathbf{W}_{hi}, \mathbf{W}_{hc}, \mathbf{W}_{ho} \in \mathbb{R}^{H^2}, \text{and} \tag{27}$$
$$\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_c, \mathbf{b}_o \in \mathbb{R}^{BxH}. \tag{28}$$

For an illustration, refer to Figure 1. The LSTM has four main gates that respond to the values of four functions determined by $\mathbf{f}, \mathbf{i}, \mathbf{c}$ and $\mathbf{o}$, represented in equations 18 through 21. With the input data matrix $\mathbf{x}_t$ (data vector if $B = 1$) concatenated with previous output matrix $\mathbf{h}_{t-1}$ (vector if $B = 1$), the flow of inputs and outputs from the various gates described in the LSTM equations interact as follows:

1. $\mathbf{h}_{t-1}$ and $\mathbf{X}_t$ are fed into the gate (or function) $\mathbf{f}$, where the output $\mathbf{f}_t$ lies in the open interval $(0, 1)$. $\mathbf{f}_t$ then interacts with previous cell state $\mathbf{c}_{t-1}$ through element-wise multiplication $\otimes$, thus $\mathbf{c}_{t-1}$ holds an interim cell state, $\mathbf{f}_t\mathbf{c}_{t-1}$. At this stage, $\mathbf{f}_t\mathbf{c}_{t-1}$ represents a state that has forgotten some previous cell state data in $\mathbf{c}_{t-1}$ that was captured as unimportant (note that importance is regulated by weight coefficients that are trained and stored in their respective weight matrices).
2. Whereas the forget gate $\mathbf{f}_t$ focuses on regulating the extent to which previous data is forgotten, the input gate $\mathbf{i}_t$ focuses on adding new data, scaled by its importance, or extent to which data should be added from the matrix comprised of $\mathbf{h}_{t-1}$ and $\mathbf{X}_t$.
3. The $tanh$ gate obtains $\mathbf{h}_{t-1}$ and $\mathbf{X}_t$, but uses the hyperbolic tangent $tanh$ function to compute its outputs (between -1 and 1).
4. The result given by $tanh$, and $\mathbf{i}_t$ are then multiplied element-wise and further added ($\bigoplus$) to $\mathbf{f}_t\mathbf{c}_{t-1}$, giving $\mathbf{c}_t$, shown in equation 20.
5. The output gate $\mathbf{o}_t$ decides what values to output, given $\mathbf{h}_{t-1}$ and $\mathbf{X}_t$, and also computes its exposure to the following cell state based on trained importance.
6. Finally, the values of the cell state, $\mathbf{c}_t$, are passed through a $tanh$ function and multiplied by the output gate result, $\mathbf{o}_t$, such that the LSTM unit keeps only the output that it accounts for as important in $\mathbf{h}_t$, described by equation 23.

2.12.1. LSTM Problem Design

Let B (different from the B above) represent a $n \times T$ matrix containing the n number of e-Behaviour sequences of all students. T is the length of all each student's e-Behaviour sequence. Let $B(s)$ represent a $1 \times T$ variable representing the e-Behaviour sequence of student s, and let $B(s)_t$ be a scalar representing the value of $B(s)$ at time t. The LSTM learns the interdependencies between variables B and $B(s)$ with the aim of classifying the risk (or determining the Safety Score) of s given the values of B and $B(s)$. That is, B and $B(s)$ are predictors of the Safety Score *(classification)* of s. Without loss of generality; this framework is used to predict the Safety Score of all students in Sections 3.1, 3.6 and 3.7.

2.13. Evaluation Metrics for Student Risk Classification

The Result Summary below is used as an evaluation template for the classification problems in Sections 3.1, 3.6 and 3.7.

Results Summary

The best-case scenario (where the e-Behaviour model obtains a 100% accuracy) occurs where all students with an At-risk Outcome label are Flagged, and all students with a Safe Outcome label receive an Ignored Safety Score.

Given an Outcome, At-risk, precision measures the proportion of students who were correctly Flagged as At-risk, a, against the total number of Flagged students. Precision is calculated as

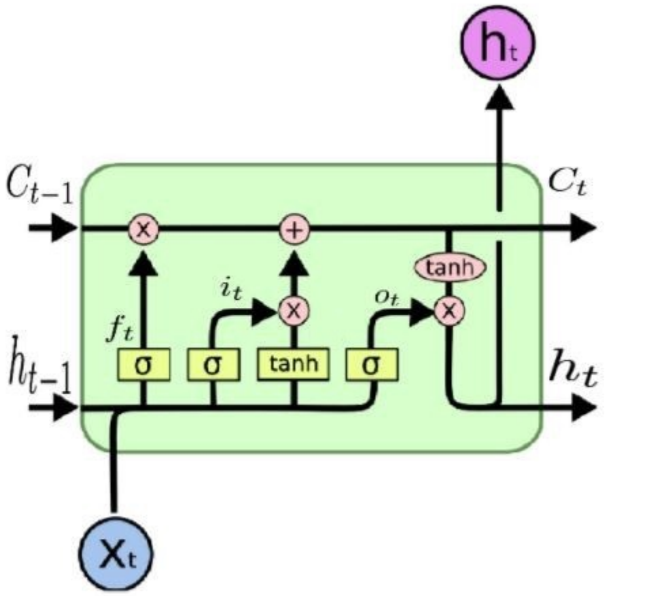

Figure 1. The LSTM unit keeps a cell state throughout its operations, which serves as input in the next time step. It also outputs $\mathbf{h}_t$, which supplements the input $\mathbf{X}_t$ in the following time step. From Olah [53].

		Safety Score (Prediction)		
		Flagged	**Ignored**	**Total**
Outcome (True Label)	**At-risk**	a	b	$a+b$
	Safe	c	d	$c+d$
	Total	$a+c$	$b+d$	$N = a+b+c+d$

Outcome	**Precision**	**Recall**
At-risk	$\frac{a}{a+c}$	$\frac{a}{a+b}$
Safe	$\frac{d}{d+b}$	$\frac{d}{d+c}$

$\frac{a}{a+c}$, where c represents the number of students who should have been Ignored as Safe. Recall measures the proportion of students who were correctly Flagged as At-risk, a, against all At-risk observations, $\frac{a}{a+b}$. The same calculation generalises to the Safe Outcome.

It is essential to know the most important metrics to measure when evaluating a classifier's performance. Consider the Results Summary above. A perfectly accurate model results in a b and c equal to 0. The precision and recall scores would be 1 for both the At-risk and Safe Outcomes. None of our models achieves perfect accuracy – they make trade-offs regarding precision and recall. For a model whose objective is to classify all students who are at risk of failing, higher precision-recall scores for the At-risk Outcome is preferred over higher precision-recall scores for the Safe Outcome. Furthermore, maximising the recall of the At-risk Outcome, $\frac{a}{a+b}$ (where the classifier recalls all students who are at risk) is preferred to maximising either the precision of the At-risk Outcome or the precision-recall scores of Safe students. Recall-maximisation will likely cause a low precision for the At-risk Outcome class (a high c). In such a case, however, no student who is At-risk will have been incorrectly Ignored.

2.13.1. The Overall Accuracy of a Model

While precision and recall are important metrics to measure a binary classifier's performance, they represent four different views of accuracy that must be analysed separately (precision and recall for At-risk and Safe Outcomes). When evaluating a model or accuracy, it is useful to obtain a single metric. A widely-used *accuracy* measure calculates the ratio between the correctly classified number of observations and the total number of observations. This *accuracy* is a good measure for balanced data, not for imbalanced data. For instance, if a test dataset contains 100 observations with an At-risk:Safe split of 10:90, a classifier can obtain an accuracy of 90% by classifying all students as Safe. An accuracy measure that combines the harmonic mean of precision and recall of either class is the *f-1 score*, whose effectiveness is surveyed by Hand and Christen [54]. The f-1 score produces two metrics (one for each Outcome) and does not concisely summarise the model's accuracy. By contrast, *Cohen's Kappa*, κ [55] is a metric that captures accuracy with a single value. The formula,

$$\kappa = (p_o - p_e)/(1 - p_e), \tag{29}$$

measures the *agreement* between the *predicted* Safety Score and the *true* Outcome. Landis and Koch [56] suggest using the scale in Table 8 to interpret the significance of κ values. In Identity 29, the observed accuracy (ratio between correctly classified number of students and total students), p_o, is adjusted for the expected *accuracy* when the classifier assigns a label randomly, p_e. In the example above, $p_e = 0.90$ and $p_o = 0.90$, giving a κ value of 0.00, or *no agreement* between the Outcomes and the Safety Scores assigned by the classifier. κ is thus more representative of a model's performance than the *accuracy* commonly used for data with balanced labels. *Chance* is an event that occurs when a classifier fails to fit an optimised objective function or has not learned anything from the data. In the above example, the κ of 0.00 signifies that the classifier performs no better than chance.

Table 8: Cohen's Kappa Interpretation

κ	Level of Agreement
< 0.00	Worse than chance
0.00 – 0.20	*Slight* agreement
0.21 – 0.40	*Fair* agreement
0.41 – 0.60	*Moderate* agreement
0.61 – 0.80	*Substantial* agreement
0.81 – 1.00	*Near-perfect* agreement

3. Results

3.1. Background Data and Grade

Figure 2 shows the linear correlation between the five Background predictor variables chosen with *Scikit-Learn*'s Recursive Feature Elimination algorithm and the Grade target variable, with a Decision Tree as its optimiser.

Quintile ($r = 0.170$) has the strongest linear correlation with Grade, followed by Township School ($r = -0.140$). The Background data's correlations suggest that higher Quintile high schools generally performed better than students from lower Quintile schools. Students from Township schools performed worse than students from other schools.

Understanding that a relationship exists between the chosen features and Grade shows that these features can inform the student's Grade and Outcome.

3.1.1. Classifying a Student based on Background Data:

The classifier used was the Decision Tree Classifier. The train-set contained 3 798, while the test-set contained 950 students. The train-test split was stratified by the Outcome of the students. A grid search on the train-set suggested a maximum tree depth of 6 and 8 maximum leaves for the Decision Tree as presented in Table 9.

If we refer to Table 9 we note that 640 out of the 950 test observations were classified correctly, producing a κ of 0.18 (slight agreement) between the Safety Score and the Outcome. The precision score for the *Flag* students suggests that 107 of the 260 Flagged students were correctly Flagged; the remaining 153 were meant to be Ignored.

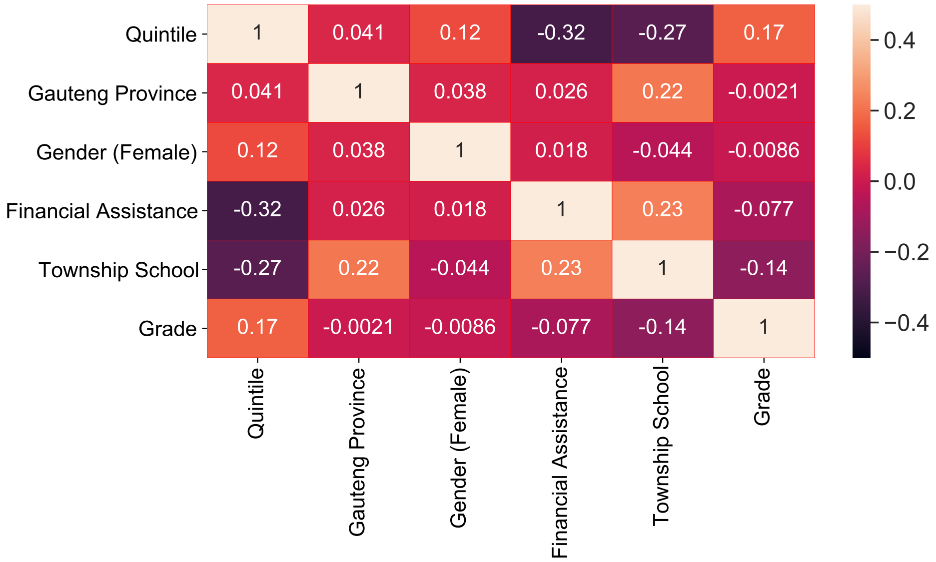

Figure 2. Pearson Correlation Coefficients of the Chosen Features. Quintile and Township School have the highest correlation with Grade

Table 9: Confusion Matrix and Summary of Background-Grade Test Set Results

		Safety Score (Prediction)		
		Flagged	**Ignored**	**Total**
Outcome (True Label)	**At-risk**	107	157	264
	Safe	153	533	686
	Total	260	690	950

Outcome	**Precision**	**Recall**
At-risk	0.41	0.41
Safe	0.77	0.78

$\kappa =$	0.18

3.2. Extraversion-levels and Grade

Although the increase in the mean Grade with Extraversion-level is apparent from the line of best-fit, this claim is confirmed by the OLS Regression model's output in Table 10. This result shows that students in higher Extraversion-levels tend to achieve higher Grades, on average. The fit described in OLS Summary Table 10 shows a linear relationship, $\hat{G}_E = 1.269E + 62.422$. The p-values of 0.000 signify that $\hat{G}_E = 1.269f + 62.422$ is not a relationship by chance. Also, the high r of 0.846 signifies that G_E moves closely with E and can be inferred from E with a 95% confidence that $\beta_0 \in [0.771, 1.767]$.

Extraversion-levels are ordinal, with each level indicating the number of posts in that level. Therefore, an E of 1 is a *lower* Extraversion-level than an E of 2. Table 13 shows the OLS Regression Summary for E against $\mathbb{E}[Gd_i]$. The linear equation of the line of best-fit is given by $\hat{G}d_i(s_i) = 0.528\mathbb{E}[Gd_i] + 35.607$. The β_0 coefficient of $\mathbb{E}[Gd_i]$ and its statistical significance ($p = 0.011$) indicates that the a marginal increase in $\mathbb{E}[Gd_i]$ of 1 Grade point corresponds to an average increase of 0.528 in $Gd_i(s_i)$. The $r(\mathbb{E}[Gd_i], Gd_i(s_i))$ of 0.421 suggests that there is a strong correlation between the mean Grade of a Discussion ($\mathbb{E}[Gd_i]$) and the Grade of a student, ($Gd_i(s_i)$), chosen at random, who participated in that Discussion. This correlation also holds for any other set of randomly-selected students.

Table 10: OLS Regression Summary – Extraversion-level Grade against Extraversion-level

Linear Equation: $\hat{G}_E = 1.269E + 62.422$

Feature	**Coeff.**	**r**	p**-value**	**Coeff. 95% CI**
E	1.269	0.846	0.000	[0.771, 1.767]
Intercept	62.422		0.000	[58.354, 66.491]

3.3. Conscientiousness and Grade

There is a positive relationship between $C(s)$ and $G(s)$, with a β_0 coefficient p-value of 0.000 as supported by Table 11. An increase of 1 in a student's Conscientiousness-level corresponds to an average increase of 5.988 Grade points out of 100.

Table 11: OLS Regression Summary – Average Number of Weekly Active Days against Grade

Linear Equation: $\hat{G}(s) = 5.988C(s) + 39.829$

Feature	**Coeff.**	**r**	p**-value**	**Coeff. 95% CI**
$C(s)$	5.988	0.319	0.000	[4.129, 7.847]
Intercept	39.829		0.000	[34.426, 45.232]

3.4. Student Discussions and Grade

The OLS Regression Summary for the linear relationship between $\mathbb{E}[Gd_i]$ and $Gd_i(s_i)$ in Table 13 shows that the linear equation of the line of best-fit is given by $\hat{G}d_i(s_i) = 0.528\mathbb{E}[Gd_i] + 35.607$. The β_0 coefficient of $\mathbb{E}[Gd_i]$ and its statistical significance (p = 0.011) indicates that the a marginal increase in $\mathbb{E}[Gd_i]$ of 1 Grade point corresponds to an average increase of 0.528 in $Gd_i(s_i)$. The $r(\mathbb{E}[Gd_i], Gd_i(s_i))$ of 0.421 suggests that there is a strong correlation between the mean Grade of a Discussion ($\mathbb{E}[Gd_i]$) and the Grade of a student, ($Gd_i(s_i)$), chosen at random, who participated in that Discussion. This correlation also holds for any other set of randomly-selected students.

Table 12: Discussion Table – Random Student's Grades against Discussion's Grade Averages

$\mathbf{d_i}$	$\mathbf{s_i}$	$\mathbb{E}[\mathbf{Gd_i}]$	$\mathbf{Gd_i(s_i)}$
0	23	83.08	92.75
1	728	56.81	59.25
2	833	49.75	48.50
⋮	⋮	⋮	⋮
336	79	75.35	90.75
337	15	74.87	70.25

3.5. Student Collaboration-groups and Grade

Table 15 shows the OLS regression results of the fit between $Gc_i(h_i)$ and $\mathbb{E}[Gc_i]$.

The coefficient of $\mathbb{E}[Gc_i]$ in Table 15 shows that the a marginal increase in $\mathbb{E}[Gc_i]$ of 1 Grade point corresponds to an estimated increase of 0.984 in $Gc_i(h_i)$. The $r(\mathbb{E}[Gc_i], Gc_i(h_i))$ of 0.479 suggests that there is a strong correlation between the average Grade of a Collaboration-group – $\mathbb{E}[Gc_i]$ – and the Grade of its Host –$Gc_i(h_i)$.

3.6. B-PM and Outcome

Table 16 shows B-PM's results. The κ of 0.51 shows a moderate agreement between B-PM's predicted Safety Scores and actual student Outcomes. B-PM's precision for the At-risk Outcome

Table 13: OLS Regression Random-Student's Grades against Discussion's Grade Averages

Linear Equation: $\hat{G}d_i(s_i) = 0.528\mathbb{E}[Gd_i] + 35.607$

Feature	**Coeff.**	**r**	**p-value**	**Coeff. 95% CI**
$\mathbb{E}[Gd_i]$	0.528	0.421	0.011	[0.131, 0.925]
Intercept	35.607		0.014	[7.789, 63.425]

Table 14: Collaboration-groups and Grades

$\mathbf{h_i}$	$\mathbf{c_i}$	$\mathbf{Gc_i(h_i)}$	$\mathbb{E}[\mathbf{Gc_i}]$
1	**{5, 48, 3, 138}**	73.00	47.68
2	**{119, 172, 199}**	81.80	67.62
3	**{40, 35, 20, 16, 51}**	90.75	69.80
4	**{90, 200, 28, 33, 94, 142, 42, 101, 84}**	49.00	62.08
5	**{81, 209, 143, 206, 12, 150}**	98.25	63.04
6	**{142, 33, 28, 42}**	59.25	58.25
7	**{65, 190, 107, 8, 173}**	46.60	74.51

group shows that out of the 39 Flagged students, 27 were Flagged correctly (since they ended up at risk of failing). 12 out of the 39 Flagged students were not meant to be Flagged. The At-risk recall indicates that out of 46 At-risk students, 27 were correctly Flagged, and the remaining 19 were incorrectly Ignored.

B-PM performed better at classifying Safe students than at classifying At-risk students: only 12 out of 124 Safe students were incorrectly Flagged, and 19 out of 131 Ignored students were wrongly Ignored.

3.7. BM and Outcome

This section reports on the results of a modified model of B-PM – without the $\{E(s)_t\}$ and $\{C(s)_t\}$ input Sequences. The comparison helps determine the change in the accuracy of B-PM after removing its personality components. This resulting model is called the Behaviour Model (BM); the only difference between BM and B-PM is that BM has only one input Sequence, $\{L(s)_t\}$.

Table 17 shows BM's results. For reference, the comparable B-PM results are shown in brackets. The At-risk recall of BM equals the At-risk recall of B-PM, meaning that BM correctly Flagged as many At-risk students as B-PM did. BM achieved a κ of 0.40.

While we only showed B-PM and BM predictions for the end of the 17 weeks, the models also produced predictions at the end of each week. Flagging students at risk earlier may be more beneficial to a student and an institution's stakeholders since early flagging allow more time for interventions – Section 4.6 reports on the trade-off between timeliness and accuracy.

4. Discussion

4.1. Background and Grade

Bourdieu and Richardson [26] argues that Cultural and Economic Capital regulate the level of success attainable by individuals. The Pearson correlation coefficients for Grade against Quintile and Township School were 0.17 and -0.14, respectively. The Decision Tree used to classify students at risk produced a κ of 0.18 (slight agreement) between the Safety Score and the Outcome. The above relationships between Background and a student's academic output provide evidence for the theories extended by Bourdieu and Richardson [26].

Table 15: OLS Regression Summary – Random-Student's Grades against Collaboration-group's Grade Averages

Linear Equation: $\hat{G}c_i(h_i) = 0.984\mathbb{E}[Gc_i] + 5.175$

Feature	**Coeff.**	**r**	p**-value**	**Coeff. 95% CI**
$\mathbb{E}[Gc_i]$	0.984	0.479	0.004	[0.334, 1.663]
Intercept	5.175		0.797	[-35.501, 0.975]

Table 16: Confusion Matrix and Summary of B-PM Test Set Results

		Safety Score (Prediction)		
		Flagged	**Ignored**	**Total**
Outcome (True Label)	**At-risk**	27	19	46
	Safe	12	112	124
	Total	39	131	**170**

Outcome	**Precision**	**Recall**
At-risk	0.69	0.59
Safe	0.85	0.90

$\kappa =$	0.51

4.2. Conscientiousness and Grade

Refer to Table 11. The β_0 coefficient of $C(s)$ indicates an increase of 1 in Conscientiousness-level is associated with an increase of 5.988 in Grade. Out of the 102 students who ended up at risk of failing their programmes (Grade < 51), 76 had a Conscientiousness-level below 3. The statistically significant positive correlation between $C(s)$ and $G(s)$ shows that $C(s)$ is a suitable predictor of a student's Outcome.

Hung and Zhang [57] presented a comparable finding; students who accessed course materials 18.5 times or more throughout their programmes obtained a grade of 77.92 out of 100 or higher, while students who accessed course materials more than 44.5 times obtained a grade of 89.62 or higher. Closely related to the above relationship is this study's findings of the correlation between a student's Extraversion-level, $E(s)$, and Grade.

4.3. Extraversion and Grade

The positive β_0 coefficient of 1.269 signifies that the average Grade of students in higher Extraversion-levels is higher than the average Grade of students in lower Extraversion-levels. While one more post than the last may not result in an additional 1.269 points to a student's Grade record, the average Grade of students who contributed to discussions more frequently, in general, was higher than the Grades of students who posted less often. Although this model accounts for the observed effect on Grade of only one independent variable, Extraversion, the probability (p-value) of Extraversion having no relationship with Grade is 0. This shows a statistical significance of Extraversion as a regressor against student Grade. An increase of 1 in the $E(s)$ correlated with an average Grade increase of 1.269. The Extraversion-Grade relationship is linked to the social science concept of *social capital* for the formation of Academic-groups.

4.4. Academic-Groups and Social Capital

The Extraversion-Grade relationship is linked to the social science concept of *social capital*, which was used as a theoretical basis for our formation of Academic-groups. Romero *et al.* [58] obtained a classification accuracy of 60% for the expected grade category of a student (Fail, Pass,

Table 17: Confusion Matrix and Summary of BM Test Set Results

		Safety Score		
		Flagged	**Ignored**	**Total**
Outcome	**At-risk**	27[27]	19[19]	46
	Safe	22 [12]	102 [112]	124
	Total	49 [39]	121 [131]	**170**

Outcome	**Precision**	**Recall**
At-risk	0.55 (0.69)	0.59 (0.59)
Safe	0.84 (0.85)	0.82 (0.90)

$\kappa =$	0.40
$(\kappa =$	$0.51)$

Good, Excellent) against her LMS behaviour. Among the features used by Romero *et al.* [58] is the number of messages sent to a forum (Extraversion-level).

Bhandari and Yasunobu [59] and other research do not illustrate the quantitative effect of social capital. However, the authors cite that 'an individual who creates and maintains social capital subsequently gains advantage from it [social capital]'. The quantification of the perceived effect of social capital was illustrated by the correlation between the Grade of a student in an Academic-group, $Gc_i(h_i)$, and the average Grade of the group, $\mathbb{E}[Gc_i]$. $Gc_i(h_i)$ responded with a statistically significant increase of 0.984 to a $\mathbb{E}[Gc_i]$ increase of 1.

Despite showing different insights and patterns, both the Discussion and Collaboration-group methods show that the *quality* of a student's academic output (Grades) is associated with the quality of academic output of her social capital. As stated in Section 1.2, a student may choose to *leverage* her social capital (that is available to all students in a cohort) by becoming part of an Academic-group.

Our Academic-group and Grade relationship, findings like Romero *et al.* [58]'s and the above statement by Bhandari and Yasunobu [59] provide evidence for the positive relationship between student success and the accumulation of social capital. This section's work has contributed to the theory that:

> *The quality of a student's social capital is the quality of her Academic group's performance.*

4.4.1. Academic-group Size Constraints:

Each Discussion and Collaboration-group were constrained to a minimum size of 3. When the sizes were reduced to 2, all linear relationships between the student's Grade and the group's average Grade collapsed and were statistically insignificant. Furthermore, residuals, $Gd_i(s_i) - \hat{G}d_i(s_i)$ (for Discussion-Grade relationships) and $Gc_i(h_i) - \hat{G}c(h)$ (for Collaboration-group-Grade relationships) are not normally distributed for the group sizes of 2.

4.5. BM and B-PM

Work from cited authors does not discuss how changes in student behaviour related to changes in student performance. Section 3.6 constructed a temporal e-Behaviour machine learning model that yielded a κ of 0.51 against a student's performance.

4.6. BM and the Trade-off Waterfall

The trade-off between the *benefit of intervention timeliness* and the *cost of intervention success* can help identify whether there are patterns over each year that can inform users about the *optimal* week (t^*) to intervene. If the current year is 2019, then t^* for 2019 (t^*_{2019}) can be determined by either one or a combination of the following factors:

1. t^*_{2019} is chosen to be the t in 2018 that yielded the maximum value of $\kappa(t)$ in 2018.

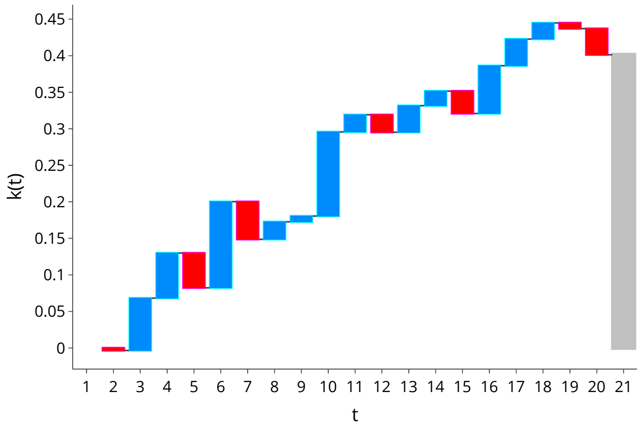

Figure 3. Trade-off Between Timeliness and Accuracy

2. t^*_{2019} is based on *exogenous* considerations determined by the institution's stakeholders. Examples of exogenous considerations are the urgency required for intervention and resources required to make interventions.

For instance, the 2018 Trade-off Waterfall was not available to this study. So the $t^*_{2019} = 17$ used in this cohort's B-PM and BM was based only on the exogenous consideration that interventions should be made by week 17.

4.6.1. Practical Benefits and Limitations of the Trade-off Waterfall

The Trade-off Waterfall is computed after the Outcome of the students is made available. It does not show the week that produces the highest accuracy in real-time, and in some weeks, the trade-off between accuracy and timeliness is not positive. For example, observe that $\kappa(15) < \kappa(14)$. In exchange for delayed intervention, BM produces a worse κ score from $\{L(s)_t\}$, which would not have made the delay worthwhile. A similar observation is made for the delay between weeks 18 and 19. Therefore, there is no way to infer the optimal week to make predictions and interventions in real-time. Instead, the Trade-off Waterfall indicates that

1. the general trade-off is that $\kappa(t)$ increases as t increases, and
2. the trade-off peaks at some point, and in this case, three weeks before the examination period at $t = 18$. Therefore, it may not be worth waiting for the start of an examination period (such as $t = 21$) before conducting interventions.[3] For example, the Trade-off Waterfall shows that after $t = 18$, there is no benefit of waiting for an extra one, two or even three weeks to intervene because $\kappa(19), \kappa(20), \kappa(21) < \kappa(18)$.

5. Limitations and Future Work

This research is a study on the methodology that guides the use of machine learning in academic performance analysis rather than the efficiency and improvement of the algorithms themselves.

Our results will likely differ across contexts since different data and algorithm configurations can generate several model outputs. The results obtained serve only as proxies for the possible outputs in academic performance research.

In the domain of LMS system user engagement, there are no formal definitions and standards, analogues or equivalent metrics that proxy a student's e-Behaviour and personality from LMS data. We modelled features as well-understood traits to further understand the relationships between behaviour, personality and academic performance.

[3] Given the login data of this cohort. Different cohorts and different datasets from those presented in this report may produce different peak-periods

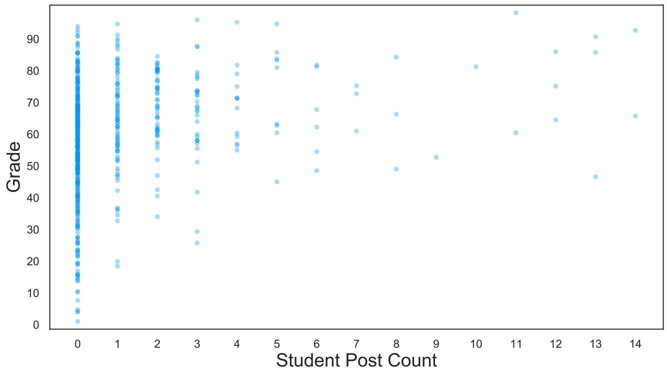

Figure 4. Crude Post Count against Student Grade. There is a positive relationship between Post Count and Grade that is not suitable for a linear OLS fit.

An unknown in all model outcomes was the presence of causality. For instance, whether e-Behaviour has an *effect* on performance is not known. Although the methodology followed aimed to set up conditions for inference, diction such as *tend to correspond with*, and *have relationships with*, instead of *causes*, showed sensitivity to all likelihood of effects from confounding variables.

In encoding performance (student Grade), we used uniform importance across all modules. We did so despite some students' modules accounting for a higher proportion of points towards obtaining a qualification from the University. We did not have access to each student's relative weighting of each module, and therefore did not account for the differences in module weightings.

5.1. Extraversion-levels

Placing the students in Extraversion-levels satisfied the OLS assumptions. While regressing each student's post count against her Grade produced statistically significant results, the data's distribution violated OLS assumptions. Hence the transformation by placing each student into an Extraversion level. Figure 4 shows the Crude Post Count against the Grade of each student. The Residual Normality, Independence, and Homoscedasticity assumptions were violated by the OLS model fitted on the data in Figure 4.

5.2. Alternative Formulations of Personalities

An approach to capture various facets of Extraversion and Conscientiousness was attempted. For instance, the *orderliness* facet of Conscientiousness required a metric that models routine or consistency of engagement. Orderliness was modelled by computing the sum of the squared deviations, SS, from each student's mean number of logins. However, the regression model that correlated Grades with SS violated the normality-of-residuals test for normality. Thus, the test for a relationship between SS and Grade was inappropriate under a linear regression model.

Personality tests, as conducted by Costa *et al.* [60], could be conducted on students in our study. Using personality assessments as an evaluation tool would help understand the extent to which the proxies we developed correspond to standard personality assessment procedures and may lead to improved proxies.

6. Conclusion

We looked at student Background, behaviour, personality, and how these factors are related to student performance. The main difference between our methodology and previous work is that we engineered features from an LMS system. We used these LMS features to act as proxy features for e-Behaviours and personality traits as input to our machine learning models. We then analysed the model outputs and their practical implications. The results demonstrated that a student's background has lower predictive power of academic performance than does her e-Behaviour and personality. We found that modelling student behaviours and personality traits require considering how accurately our proposed e-Behaviour and personality proxies model *true* behaviours and personality traits – we based our models on definitions found in previous literature.

We were able to use Bourdieu's Three Forms of Capital to model social, economic and cultural capital, and the Big Five personality traits to model e-Behaviour and personality. From student Background and LMS forum engagement data, Bourdieu's Three Forms of Capital were modelled in the following ways:

- Economic Capital – modelled by the *Financial Assistance*,
- Cultural Capital – modelled by the *Quintile in Province* and *Township School*, and
- Social Capital – modelled by Academic-groups.

The correlation values for Financial Assistance, Quintile and Township School provides evidence for the authors' argument that Cultural, Economic and Social Capital regulate the level of success attainable by individuals. Cultural and Economic Capital, combined with the Gender feature, performed better than chance at predicting student performance. A student's quality of Social Capital available to her also correlates positively with her academic performance.

We used two of the Big Five personality traits, Conscientiousness and Extraversion, previously found to correlate strongly with performance. Conscientiousness and Extraversion each showed significant predictive performance when used in our MNL models. With these personality features, our e-Behaviour classifier achieved better accuracy than without the personality features. The works cited do not discuss how changes in student behaviour relate to changes in their performance. We constructed a temporal e-Behaviour model using deep learning that showed an increase in accuracy over time. This e-Behaviour learning can be used to flag students at risk at any point throughout their study programmes.

The analyses in this research are practically useful if they can inform or influence student behaviour. Firstly, a student may find helpful the linear relationship between academic performance and Extraversion. The empirical evidence that a higher Extraversion level is associated with a better academic performance may encourage students to engage in forums more frequently. This evidence may encourage her to engage with the academic content more thoroughly to contribute meaningfully to discussions. Secondly, we showed, using unsupervised cluster learning, that a student's performance is congruent with the performance of her Academic-group. The above result is a further motive to action a student into leveraging her social capital by engaging in forums more frequently, since engagement increases her chances of being in an Academic-group.

References

1. Richiţeanu-Năstase, E.R.; Stăiculescu, C. University dropout. Causes and solution **2018**. *1*, 71–75.
2. Wright, D.; Taylor, A. Introducing Psychology: An Experimental Approach, 1970.
3. Heppner, P.P.; Wampold, B.E.; Owen, J.; Wang, K.T.; Thompson, M.N. *Research design in counseling*, fourth edition ed.; Cengage Learning, 2015.
4. Stone, A.A. *The science of self-report: implications for research and practice*; Lawrence Erlbaum, 2000.
5. Northrup, D.A.; York University (Toronto, O.; for Social Research, I. *The problem of the self-report in survey research: working paper*; Institute for Social Research, York University, 1997.
6. Fellegi, I.P. The Evaluation of the Accuracy of Survey Results: Some Canadian Experiences. *International Statistical Review / Revue Internationale de Statistique* **1973**, *41*, 1–14.
7. Poropat, A.E. A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin* **2009**, *135*, 322.
8. Morris, P.E.; Fritz, C.O. Conscientiousness and procrastination predict academic coursework marks rather than examination performance. *Learning and Individual Differences* **2015**, *39*, 193–198.
9. Kim, S.; Fernandez, S.; Terrier, L. Procrastination, personality traits, and academic performance: When active and passive procrastination tell a different story. *Personality and Individual Differences* **2017**, *108*, 154–157.
10. Costa, P.T.; McCrae, R.R. *The NEO personality inventory*; Psychological Assessment Resources Odessa, FL, 1985.
11. Furnham, A.; Nuygards, S.; Chamorro-Premuzic, T. Personality, assessment methods and academic performance. *Instructional Science* **2013**, *41*, 975–987.
12. Ciorbea, I.; Pasarica, F. The study of the relationship between personality and academic performance. *Procedia-Social and Behavioral Sciences* **2013**, *78*, 400–404.
13. Kumari, B. The correlation of Personality Traits and Academic performance: A review of literature. *IOSR Journal of Humanities and Social Science* **2014**, *19*, 15–18.
14. Chamorro-Premuzic, T.; Furnham, A. Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of research in personality* **2003**, *37*, 319–338.
15. Costa Jr, P.T.; McCrae, R.R. *The Revised NEO Personality Inventory (NEO-PI-R).*; Sage Publications, Inc, 2008.
16. Wilt, J.; Revelle, W., Extraversion; 2009; pp. 27–45.

17. Association, A.P. APA Dictionary of Psychology – Gregariousness. https://dictionary.apa.org/gregariousness, 2020. Accessed: 2020-12-29.
18. Merriam-Webster. Dutiful. https://www.merriam-webster.com/dictionary/dutifulness. Accessed: 2020-11-29.
19. Akçapınar, G. Profiling students' approaches to learning through moodle logs. Multidisciplinary Academic Conference on Education, Teaching and Learning (MAC-ETL 2015), 2015.
20. Huang, A.Y.; Lu, O.H.; Huang, J.C.; Yin, C.J.; Yang, S.J. Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs. *Interactive Learning Environments* **2020**, *28*, 206–230.
21. Khan, I.A.; Brinkman, W.P.; Fine, N.; Hierons, R.M. Measuring personality from keyboard and mouse use. Proceedings of the 15th European conference on Cognitive ergonomics: the ergonomics of cool interaction, 2008, pp. 1–8.
22. Fowler, G.C.; Glorfeld, L.W. Predicting aptitude in introductory computing: A classification model. *AEDS Journal* **1981**, *14*, 96–109.
23. Poh, N.; Smythe, I. To what extend can we predict students' performance? A case study in colleges in South Africa. 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 2014, pp. 416–421. doi:10.1109/CIDM.2014.7008698.
24. Evans, G.E.; Simkin, M.G. What Best Predicts Computer Proficiency? *Commun. ACM* **1989**, *32*, 1322–1327. doi:10.1145/68814.68817.
25. Dauter, L.G. Economic sociology. https://www.britannica.com/topic/economic-sociology, 2016. Accessed: 2020-12-29.
26. Bourdieu, P.; Richardson, J.G. The forms of capital **1986**.
27. Carpiano, R.M. Toward a neighborhood resource-based theory of social capital for health: Can Bourdieu and sociology help? *Social science & medicine* **2006**, *62*, 165–175.
28. Hallinan, M.T.; Smith, S.S. Classroom characteristics and student friendship cliques. *Social forces* **1989**, *67*, 898–919.
29. Song, L. Social capital and psychological distress. *Journal of health and social behavior* **2011**, *52*, 478–492.
30. Hayes, E. Elaine Hayes on "The Forms of Capital", 1997.
31. Smith, E.; White, P. What makes a successful undergraduate? The relationship between student characteristics, degree subject and academic success at university. *British Educational Research Journal* **2015**, *41*, 686–708.
32. Caldas, S.J.; Bankston, C. Effect of School Population Socioeconomic Status on Individual Academic Achievement. *The Journal of Educational Research* **1997**, *90*, 269–277, [https://doi.org/10.1080/00220671.1997.10544583]. doi:10.1080/00220671.1997.10544583.
33. Fan, J. The Impact of Economic Capital, Social Capital and Cultural Capital: Chinese Families' Access to Educational Resources. *Sociology Mind* **2014**, *04*, 272–281. doi:10.4236/sm.2014.44028.
34. Ajzen, I. *Attitudes, personality, and behavior*; McGraw-Hill Education (UK), 2005.
35. Campbell, D.T., Social Attitudes and Other Acquired Behavioral Dispositions. In *Psychology: A study of a science. Study II. Empirical substructure and relations with other sciences. Volume 6. Investigations of man as socius: Their place in psychology and the social sciences.*; McGraw-Hill, 1963; p. 94–172. doi:10.1037/10590-003.
36. Hemakumara, G.; Ruslan, R. Spatial Behaviour Modelling of Unauthorised Housing in Colombo, Sri Lanka **2018**. *25*, 91–107. doi:10.21315/kajh2018.25.2.5.
37. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Machine learning* **2002**, *46*, 389–422.
38. Barrick, M.R.; Mount, M.K.; Strauss, J.P. Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of applied psychology* **1993**, *78*, 715.
39. Campbell, J.P. Modeling the performance prediction problem in industrial and organizational psychology. **1990**.
40. Stigler, S.M. Gauss and the Invention of Least Squares. *The Annals of Statistics* **1981**, *9*, 465–474.
41. Blumberg, M.; Pringle, C.D. The missing opportunity in organizational research: Some implications for a theory of work performance. *Academy of management Review* **1982**, *7*, 560–569.
42. Leontjeva, A.; Kuzovkin, I. Combining Static and Dynamic Features for Multivariate Sequence Classification. 2016, pp. 21–30. doi:10.1109/DSAA.2016.10.
43. Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics* **1991**, *21*, 660–674.
44. Mitchell, T.M. *Machine Learning*, 1 ed.; McGraw-Hill, Inc.: New York, NY, USA, 1997; chapter 10.
45. Raileanu, L.; Stoffel, K. Theoretical Comparison between the Gini Index and Information Gain Criteria. *Annals of Mathematics and Artificial Intelligence* **2004**, *41*, 77–93.
46. Khalaf, A.; Hashim, A.; Akeel, W. Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis. *International Journal of Interactive Multimedia and Artificial Intelligence* **2018**, *5*, 26–31.
47. Topîrceanu, A.; Grosseck, G. Decision tree learning used for the classification of student archetypes in online courses. *Procedia Computer Science* **2017**, *112*, 51 – 60. Knowledge-Based and Intelligent Information Engineering Systems: Proceedings of the 21st International Conference, KES-20176-8 September 2017, Marseille, France.
48. Kolo, K.D.; Adepoju, S.A.; Alhassan, J.K. A decision tree approach for predicting students academic performance. *International Journal of Education and Management Engineering* **2015**, *5*, 12.
49. Gujarati, D.N.; Porter, D.C. *Basic Econometrics*; Douglas Reiner, 2009.
50. Liu, Z.; Sullivan, C.J. Prediction of weather induced background radiation fluctuation with recurrent neural networks. *Radiation Physics and Chemistry* **2019**, *155*, 275 – 280.
51. Wang, M.; Zhang, Y.D.; Cui, G. Human motion recognition exploiting radar with stacked recurrent neural network. *Digital Signal Processing* **2019**, *87*, 125 – 131.
52. Bengio, Y.; Boulanger-Lewandowski, N.; Pascanu, R. Advances in Optimizing Recurrent Networks. *CoRR* **2012**, *abs/1212.0901*, [1212.0901].

53. Olah, C. Understanding LSTM Networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/, 2015. Accessed: 2019-04-19.
54. Hand, D.; Christen, P. A note on using the F-measure for evaluating record linkage algorithms. *Statistics and Computing* **2018**, *28*, 539–547.
55. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* **1960**, *20*, 37–46, [https://doi.org/10.1177/001316446002000104]. doi:10.1177/001316446002000104.
56. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *biometrics* **1977**, pp. 159–174.
57. Hung, J.L.; Zhang, K. Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching* **2009**, *4*.
58. Romero, C.; Ventura, S.; Espejo, P.; Martínez, C. Data Mining Algorithms to Classify Students. 2008, pp. 8–17.
59. Bhandari, H.; Yasunobu, K. What Is Social Capital? A Comprehensive Review of the Concept. *Asian Journal of Social Science* **2009**, *37*, 480–510. doi:10.1163/156853109X436847.
60. Costa, P.; McCrae, R.; Kay, G. Persons, Places, and Personality: Career Assessment Using the Revised NEO Personality Inventory. *Journal of Career Assessment - J CAREER ASSESSMENT* **1995**, *3*, 123–139.