Open Education Analytics

Use Case Documentation: Predicting Student Well-Being Package with Tasmania Department of Education

# 

Use Case Documentation is intended to help education organization plan specific data and AI projects.

To ensure the appropriate and ethical use of data, OEA recommends applying Microsoft's Responsible AI principles. This document shows how these principles can be operationalized in the case of predicting student well-being. The principles include fairness, reliability and safety, privacy and security, inclusion, transparency, and accountability. For more information see: [Responsible AI | Open Education Analytics](https://openeducationanalytics.org/responsible-ai/).

# 1) The Use Case Problem

**Defining the Problem: What problem does this use case seek to solve?**

The Predicting Student Well-Being package is designed to help predict students in need of personalized support using student data. It shows how to develop a predictive model using key indicators related to student well-being. This package was developed in partnership with the Department of Education from Tasmania, Australia.

With this project in Tasmania, the Tasmania Department of Education and Microsoft Education collaborated under a special research agreement to develop a machine learning model to help nominate students for well-being assessment and support in Tasmanian schools. All students have different levels of need, and some require more personalized care yet the process of identifying them tends to be a subjective and case-by-case process across schools. Often, interventions are put in place only after their context has become critical.

The goal is that the model will act to augment and improve current processes, leveraging the scale of available data, without omitting humans from the core decision making around validating model outputs and designing the right interventions based on results. The sooner that issues are identified, the sooner supports can be provided.

The general approach for the modeling work is to follow through responsible AI best practices. Based on the Tasmania data, we built a human-in-the-loop framework that optimizes the model towards transparency, fairness, and robustness on top of accuracy performance. The human-in-the-loop approach contains 3 stages: Model Development, User Consumption and Evaluation, and Model Recalibration. The loop starts with a predicative model to nominate students in need of personalized support and refer them to human experts for well-being assessment and support. This way, we can include human experts early in the loop, and subsequently the human evaluation can be used to augment existing datasets and further retrain the model. In addition, the model generates explanations for why a student should be nominated in terms of the indicators collected from student profiles. These suggestions help the human experts generate insights and actionable interventions for the student population and individual students. This knowledge can also spark larger discussions and investigations on how to best support students based on their specific needs during the iterative process.

We recognize that the model built in the package is optimized with respect to the Tasmania dataset, and therefore subject to the same limitation on data quality and knowledge. As a result, when applying this package to build their own end systems, users should understand that the final performance in general is a function of the data quality, assumptions made, and the modeling quality, on top of other factors. To obtain the best results, we strongly recommend users to follow and be inspired by the best practice shown in the package:

* investigating model errors,
* generating model explanations,
* assessing model fairness, and
* creating robustness with casual inference tools

These practices are supported by the full-suite Responsible AI Toolbox and explanation framework [SHAP](https://github.com/slundberg/shap) offered by Microsoft.

Recommendations on Responsible AI System Design

In developing a well-being support system for students, we recommend users of this package follow or be inspired by this three-stage human-in-the-loop approach for responsible AI best practice:

**Model Development**: This stage is where the package can be used to develop a predictive model that adapts to specific student datasets for an end-goal defined by the user. Through iterations of model design, we find the best use of the Tasmania dataset is to build a model to proactively nominate students in need of personalized support and then refer them to human experts for well-being assessment and support. Inside the package, the data engineering pipeline demonstrates how to go from pseudonymized dataset to features ready for predictive model consumption – a process called featurization in the data science world; the ML model pipeline demonstrates how to use the features to generate predictions and explanations. These predictions represent the probabilities that a student should be nominated for further well-being assessments, while the explanations represent what features make the model come to such predictions. In addition, we demonstrate model monitoring, error analysis, fairness assessment, and casual inference to enhance the robustness of the model. These analyses can gather useful insights into the model and surface important questions about the modeling process to human experts in the next stage.

**User Consumption and Evaluation**: This stage is where the model output aid different users in the educational system. Starting from the top level, school administrators can monitor trends and patterns in what categories of features drive students’ need of personalized support at the aggregate level. They can also discuss how to guide resources more efficiently and create policies from this process. Subject matter experts on student well-being can assess the nominated students who potentially need personalized support and make recommendations for actions based on their specific needs. Well-being support team can use the expert evaluation results along with model insights to best support them. It is recommended that feedback from all types of users be documented, and relevant insights communicated back to the model developers in the next stage.

**Model recalibration**: This stage is where model developers calibrate the model with given human feedback. At this stage, data scientists can partner with subject matter experts and other users to incorporate their knowledge for model validation and debugging. Potential analyses can range from identifying biases in the dataset to validation of predictive performance results and top important features. Using the human feedback gathered, data scientists can then fine-tune the model by for example, addressing the biases in existing datasets by further data or feature engineering, and augmenting existing datasets or creating new ones based on the human ground truth to retrain or augment the model. If human interventions were made, changes in the outcome (nomination prediction) can also be observed and used to create more robust conclusions about the interventions using explicit casual inference approaches. These efforts can then power an improved model in the next iteration of the loop.

Human-in-the-loop System Design

Model Development

User Consumption

And Evaluation

Reiteration

Model Recalibration

# 2) Stakeholder Involvement

Considering the desired value or benefits of a use case and its potential harms requires the consideration of different stakeholders and their points of view. Stakeholders typically include the people who are responsible for, will use, or will be affected by the use case.

|  |  |  |
| --- | --- | --- |
| **Stakeholder Groups** | **Relationship to Use Case** | **Involvement in Use Case** |
| Students | Students will not directly interact with the system. Teachers and school staff will use data insights to make independent decisions regarding assessing and supporting student well-being. | Students do not interact within the systems apart from providing data that is gathered through normal schooling processes, but are the intended beneficiaries of this use case. |
| Parents or Guardians | Indirect Stakeholders | No interaction with system; child receives support from school system |
| Educators (Faculty or Teachers) | Direct stakeholders | Recipient of the insights of the model, with the opportunity to turn them into actions. |
| School or Department Leaders | Direct stakeholders | Recipient of the insights of the model, with the opportunity to turn them into actions. |
| School System or Institutional Leaders | Direct stakeholders | Take advantage of system wide insights, to inform policy decision and resourcing. |
| Researchers | Indirect stakeholders | Insights from education literature used to guide modeling decisions |
| Potential Malicious Actors | Direct stakeholders | No intended interaction; may attempt to break or hack system in ill will. |

**What is the plan for engaging these stakeholders in the use case development process?**

Including stakeholders in the early thinking and conceptualization of a data use case is a good way to ensure that the use case output will be accepted, trusted, and used by key stakeholders.

Key stakeholders from the education system-level and school level support teams for students share their perspectives on describing the process for assessing and supporting student well-being. Discussion of what data is needed in the modelling work to achieve accurate predictions, as well as what decisions can be made based on the insights from the model, should take place during these early stages.

# 3) Mapping Theory to Data

**For this use case, what prior research or conceptual model frames your theory of the problem?**

For most common education use cases, research has already been conducted or a theory of the problem developed. For example, extensive research has identified key data elements that are related to the well-being of students. This type of research, theory or model should help identify the most relevant data sources for a specific use case.

* [3-Child-and-Family-Wellbeing-Assessment-Tool.pdf](https://microsoft.sharepoint-df.com/teams/Internal-OEATeam/Shared%20Documents/4.%20Data%20Science/MAIDAP%20H2%20FY22%20-%20Tasmania/3-Child-and-Family-Wellbeing-Assessment-Tool.pdf%20(https:/www.strongfamiliessafekids.tas.gov.au/__data/assets/pdf_file/0016/5551/3-Child-and-Family-Wellbeing-Assessment-Tool.pdf)) 
  + Well-being:
    - Being loved and safe
    - Having material basics
    - Being healthy
    - Learning
    - Participating
    - Having a positive sense of culture and identity
* [Vulnerable learners in the age of COVID-19: A scoping review | SpringerLink](https://link.springer.com/article/10.1007/s13384-020-00409-5)
  + Index of Community Socio-Educational Advantage (ICSEA) for each school
  + Attendance to video lectures – indicator of receiving proper cognitive/emotional interactions (regardless of grade)
  + Students from disadvantaged backgrounds reported as being more likely to experience markers of disengagement – ex. Daily absence, disruptive behavior, poor school connectedness
  + NESCO highlights the importance of addressing the psychological challenges associated with the pandemic and recommends that this take priority over teaching. Necessity to “ensure regular human interactions, enable social caring measures, and address possible psychosocial challenges that students may face when they are isolated”
  + Australian Digital Index (Thomas et al, 2019) measures digital inclusion in 3 discrete ways: access, affordability, digital ability. Digital divide between students from low and high socio-educational backgrounds
* [Predictors of High School Graduation [36474].pdf](file:///C:/Users/jehon/AppData/Local/Packages/Microsoft.Office.OneNote_8wekyb3d8bbwe/LocalState/EmbeddedFileFolder/0/Predictors%20of%20High%20School%20Graduation%5b36474%5d.pdf)
  + Demographic characteristics, family background, prior school performance, psychological characteristics, school or community characteristics (student vs institutional)
  + Individual factors – background (demographics, health, prior performance, past experiences); attitudes (goals, values, self-perceptions); behaviors (engagement, coursework, deviance, peers, employment); performance (achievement, persistence, attainment)
  + Institutional factors – families (structure, resources, practices); schools (composition, structure, resources, practices); communities (composition, resources)

Other related links

* [Race, Gender, and Measures of Success in Engineering Education - Ohland - 2011 - Journal of Engineering Education - Wiley Online Library](https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2168-9830.2011.tb00012.x)
* [Models for early prediction of at-risk students in a course using standards-based grading - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S0360131516301634)
* [Student Attendance and Educational Outcomes: Every Day Counts](https://www.telethonkids.org.au/globalassets/media/documents/research-topics/student-attendance-and-educational-outcomes-2015.pdf)
  + Research shows that 'unauthorized absence’ has a negative impact on Educational Outcome

**Mapping theory to data and developing the ‘data dictionary.’**

A key part of the use case development process is deciding which data to use and how it should be mapped to the theory of the problem. **Please see “Privacy and Security” section below for more ensuring that sensitive data is protected.**

|  |  |
| --- | --- |
| **Theoretical Construct** | **Local Data Source Mapped to Theoretical Construct** |
| Attendance – Prior research shows that unauthorized absence is an indicator of other risk factors | * Created features represented by the minimum, average and maximum number of absent streak and present streak |
| Assessment scores – Assessment scores are reflective learning. | * Select assessment scores of certain subjects based on data quality (proportion of missing value, whether an assessment of a subject is mandatory / optional, consistency of assessment criterion). * Create features represented by the earliest assessment score (“ActualScore” attribute in the table) and latest assessment score for each student. |
| Medical Conditions – Represents the health context of a student. | * Create features standing for seriousness of medical conditions * Create features standing for whether the student used to have medical condition that raised alert. * Seriousness of certain diseases and indications of whether a student has certain type of medical alert in the records. |
| Protection Order – Represents physical well-being. | * Filter records by the date. Keep those within the timeframe wanted. * Count the number of protection orders of each individual. |
| Disability – Represents additional health context. | * Create indicators of whether a student used to have certain disability registered in the system. |
| Disciplinary –Students with disciplinary sanctions | * Flag on whether a student had certain disciplinary sanctions * Count the number of days a student was under a certain type of sanction |
| Observed (negative) behaviors represented the number of negative behaviors a student had in a specific timeframe. | * Count number of negative behaviors recorded of a student during a specific timeframe. * Count actions taken against the student during a specific timeframe. |

**What are the constraints of these datasets for this specific use case?** In most systems, the data used for modelling is not of consistent quality or representativeness. It is important to clarify and describe these weaknesses in the data. The following provides examples of constraints from the predictive modelling in the case of the Tasmania Department of Education.

|  |  |
| --- | --- |
| **Dataset Name** | **Constraints** |
| Student Enrolment | * Some students don’t have enrolment records after grade 6. It might be because they transferred from public school to private school so there is no record in the system for them. It can also be that a student drops out and doesn’t enroll in any school anymore. At current stage it’s impossible to tell if students who don’t have enrolment records are just dropping out of school or transferring to private school. * Some students have duplicated records (same info) |
| Student Assessment | * The assessment records of some students are not complete – records are missing for certain reasons, for example, some students don’t have KDC scores (an assessment that should be satisfied if she/he needs to enter grade 1) because they enter the system after graduating from kindergarten. * Complementary information needed – assessments of some subjects only need to be taken on certain student populations (e.g. TCE are required only for students at grade 11 and grade 12; ACF-MA is an optional assessment for students at grade 10, etc.) All these require efforts to clarify and clean. * Unclear and inconsistent categorization of assessments / subjects – when doing EDA we found the categorization of assessments is not consistent. Some assessments can be grouped by “AssessmentTierCode1” but some should be grouped by “AssessmentTierCode3”. * Student attendance rates were significantly affected by the COVID-19 Lockdown in Tasmania from 16 March – 2020 to Tuesday 9 June 2020. The data from this period is not comparable with the same period in previous years. * If data is complete, we may also use the normalized percentile across the grade / class or other units. |
| Student Attendance | * There are missing attendance dates for some students for various years. Some students had very little attendance data while others had significantly more data. * Student attendance rates were significantly affected by the COVID-19 Lockdown in Tasmania from 16 March – 2020 to Tuesday 9 June 2020. The data from this period is not comparable with the same period in previous years. * Tasmania moved from 3 Terms to 4 Terms – this occurred at the beginning of 2013. Previous to 2013 we had 3 Terms in a year – this Term 1 being between 15 to 16 weeks – though Term 1 incorporated a break of about 6 school days for Easter. * The dates of the Easter Holiday also effect the number of weeks in Term 1 and 2. |
| Student Protection Order | * Typos in protection order type of each record make it hard to categorize / aggregate each protection order type. * Erroneous information – multiple duplications / extremely long-lasting time of a protection order (e.g., till year 9999) * Complementary information needed – there are multiple protection orders issued at the same time on the same student. |
| Student Vulnerability Indicators (Vulnerability and LevelofNeed columns) | * The binary vulnerability indicator refers to whether a student needs personalized support and equivalently level 3 or 4 in the level of need indicator. * The current indicator in case management system contains only students who were assessed for well-being to some degree. Most students were not assessed and do not have data in the data set. * Time when students were assessed is missing (to be more specific, the time each level of need assessment was made or changed). Thus, we cannot determine the appropriate time range to be used for relevant features. * Since the exact time of when the vulnerability label was created was unknown and an approximate timeline of the labeling from April 2020 to August 2021 was assumed, efforts were made to ensure that features used in the model were dated before April 2020, the presumed first month of the labeling period, in order to avoid information leakage (i.e. using the future to predict the past). |
| (Additional information needed) Organization Identifier / School information | * Having the information about the organization of each student/ each instructor would be very helpful to tell the variousness of teachers’ perspective of well-being. * It can help with generating the feature describing “distance from school” of each student. * It will enable building of more specific models, and allow for more careful examination of the model results (before being applied on unidentified students). * However, this information was not provided and therefore not used in modeling. |

4) Responsible AI Principles Applied

Fairness Principle

|  |
| --- |
| AI systems should treat everyone in a fair and balanced manner and not affect similarly situated groups of people in different ways. To ensure AI models are trained in a way that does not embed or re-enforce human biases, models must be tested for fairness. Microsoft has developed an open-source toolkit to support this called [Fairlearn](https://fairlearn.org/), which can be applied within the Azure analytical services used in the OEA reference architecture.  **Who is most likely to be at risk of experiencing harms from this use case? For example, would any groups (immigrants, rural students) or subpopulations (gender, language group) face adverse consequences from the AI?**  Students who are predicted to have low well-being incorrectly OR students who are are in need of personalized supports but are not identified by the predictive model.  The primary risks may be related to biases inherent in the training data that is based on the subjective decisions of educators and support staff in the case of Tasmania who performed assessments of student well-being. Some of the ways in which this training data may be biased:   * Identify more students from low-income backgrounds or low-income schools as needing supports for well-being and miss students from high income backgrounds * Similar biases for gender, ethnicity, parental education level, or other demographic or behavioral variables.   **Planned Mitigations:**  Modeling pipelines include the use of Fairlearn to assess fairness across sensitive demographic subgroups in terms of multiple fairness metrics such as disparities in recall rates, accuracy rates, etc. If a particular subgroup is found to be receiving disparate harm risk, we recommend users first discuss with subject matter experts to find out the underlying causes of the disparities. The potential causes could come from the data, which may be collected under the influence of human judgment or specific selection criteria for the purpose of educational interventions.  Such was the case in the Tasmania dataset, where economically disadvantaged students were nominated much more often than their counterparts, and therefore the model has learned such a pattern and received a higher recall rate than the counterpart. Similar patterns were in students with indigenous status or independent status, and female students. Given the knowledge that the Tasmania DoE provided a framework for evaluating students’ well-being and organized support teams to nominate students according to the framework. It may be possible that the nomination process was intentionally targeting these sensitive subgroups. Further discussion among model developers and subject matter experts is therefore highly recommended to address whether this was indeed the intention, whether such disparity should be corrected, and if so, what kind of mitigation is needed, for example, a change in nomination procedure, or a change in the model as supported by the [Mitigation section](https://fairlearn.org/main/user_guide/mitigation.html) of the Fairlearn package. |
| **Are these groups and subpopulations clearly labelled in the dataset?**  Demographic information is provided in the student dataset. Additionally, viewing the impact of the model across these subpopulations is easily visible with these fairness tools. As mentioned, demographic information such as gender, economic disadvantage status, indigenous status, independent status, and birth months, were included in the Tasmania dataset. |

Reliability and Safety Principle

Systems should operate reliably and safely when they function in the world. AI systems must be designed with a view to the potential benefits and risks to different stakeholders and undergo rigorous testing to ensure they respond safely to unanticipated situations and do not evolve in ways that are inconsistent with the original shared purpose.

**What are possible risks faced by learners or educators from the analytics of this use case?**

1. Risk 1: Data drift in student populations caused by a changing current event landscape. Events such as covid can accelerate the change in model validity as new indicators for well-being become more relevant or the population faces new risks.
2. Risk 2: Data drift in student populations caused by a change in the nomination or assessment procedure. The Tasmania dataset assumed a certain nomination and assessment procedure where students are nominated by school support teams and then assessed by subject matter experts according to specific well-being criteria. And models were built in the package based on this assumption. If there were, for example, no nomination process involved in a new dataset, the models should be re-developed by the user.
3. Risk 3: Automation of well-being labeling, or over reliance on model predictions and only minimal oversight on the model.

**Planned Mitigations:**

Mandatory human-in-the-loop component for model decisions as outlined in the Model Calibration stage of the recommended System Design. At no time should stakeholders rely solely on the output of the AI system to make decisions about students’ well-being, broadly construed.

Periodic retraining of the model in relevant intervals, in which older data is excluded and new data is prioritized. Triggering retraining or recalibration can be considered based on different quantitative metrics such as:

* + Based on regular scheduled intervals (say every 6 months)
  + Based on amount of new labels updated and validated by humans
  + Based on the amount of errors detected by humans in the loop (regular validation of model decisions)

Transparency Principle

Transparency requires visibility into all levels of decision-making and design of an AI system. Designers should clearly document their goals, definitions, and design choices, and any assumptions they have made. Those who build and use AI systems should be forthcoming about when, why, and how they choose to build and deploy them, as well as their data and systems’ limitations. Information should be readily available on the quality of the predictions and recommendations the AI system makes. Transparency also encompasses intelligibility, which means that people (in this case, educators, parents, students, etc.) should be able to understand, monitor, and respond to the technical behavior or recommendations of AI systems.

[Video](https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6) on Transparency Principle

**What steps will the analytics or AI process include?**

Tasmania and Microsoft provided documentation describing the capabilities and limitations of the AI system, including, but not limited to, warnings to the end-user about relying on outputs made by the system and descriptions of the intended uses of the system. Here, the goal is to encourage stakeholders to use the system only in the ways in which it was designed and intended to be used, by clarifying exactly what those use-cases are and how the system is designed to enable them.

In terms of model interpretability – all models built are paired with existing explanation AI methods for [InterpretML](https://interpret.ml/) and [SHAP](https://github.com/slundberg/shap). These should be discussed at length with any stakeholders as well as have their limitations considered. Additionally, for some model types created with the Tasmania data pipelines, it may be possible to provide explanation details directly for each inference call. In these cases (such as explainable boosting machines), educators may be able to see a clear indication of the impact of a given feature on model results.

**The AI process includes the following:**

1. *Data exploration & cleaning*

During the data exploration phase, the analysis below was conducted for datasets mentioned previously.

1. Basic description of raw data sources
   1. Calculate average, max, min, counts, missing values, (unique values of categorical attributes), percentiles, variance, etc.
   2. Plot out distribution (box plot / histogram / violine plot / bar plot)
2. Basic data cleaning:
   1. Remove records with missing values, or impute missing data
   2. Remove outliers (extreme values) or erroneous values
   3. Remove irrelevant values
   4. Get rid of duplicate values
3. Conduct clustering or segmentation (for example, some categories have rare values within an attribute. Those can be clustered as one category to reduce the dimensionality of data.)
4. Convert data types of certain attributes (e.g., convert float to integer, replace strings with number)
5. Select attributes of interest for further investigation and feature engineering. Several criteria can be applied here, for instance, we always want to keep those attributes without too many missing values. In our specific use case, it’s important to identify attributes supported by theory for our further investigation.
6. Correlation analysis – investigate the mutual relationship or association among several attributes. For example, plot out distributions of certain attributes (e.g., KDC assessment scores) of each level of need group.
7. Join tables for creating full profile for each entity.
8. *Feature engineering*

Transforming raw data into features described in the data dictionary section above. The intention is to determine the underlying contributing factors of students’ well-being and needs for personalized supports, resulting in improved model accuracy on unseen data.

(e.g. Changing daily attendance data into ‘unauthorised absence’ and determining metrics regarding student absent/present streak)

Other typical methods:

* one-hot encoding
* binning
* clustering and aggregation

1. Model Metric selection

The primary metric used for evaluation of binary model includes accuracy and recall for fixed decision threshold, as well as AUC ROC score and precision-recall curve for soft decision thresholds:

* Accuracy: Number of correctly predictions / Number of all students.
* Recall: Number of true positives divided by the number of true positives plus the number of false negatives.
* AUC ROC score: The area under the ROC curve (AUC) is a useful tool for evaluating the quality of class separation for soft classifiers.
* Precision-recall curve: We can visualize the performance of classification models in terms of their precision-recall tradeoffs.

The recall was a specific area of focus due to emphasis on identifying all students potentially in need of personalized supports. With recall, the model training will focus on making sure all students are identified at the expense of a higher false positive rate (based on joint discussion with Tasmania project leads).

1. Model building

* The following models were used for initial experimentation: Logistic Regression, Explainable Boosting Machine (EBM), LightGBM.
* We include binary classification models (to predict students as in need or not) in the pipeline.

1. Validation and analysis of results

The dataset was split into train, validation, and test dataset. The model was first trained using the training set. Upon training the model, the performance of each model was evaluated on validation sets in a k-fold stratified cross-validation fashion. Then the performance of the model was evaluated on a test set. The model with the best performance will be selected as the final model.

**Who will develop the analytics or models?**

A small group of data scientists from Microsoft and the Tasmania Department of Education worked together to develop the model iteratively. The model was enhanced as more datasets representing more aspects of the well-being construct were added into the system.

**How will the limitations of the analytics or AI model be communicated to stakeholders and users?**

At the model building stage, there were ongoing conversations between all direct stakeholders and these conversations delved into the data limitations of the datasets used in the model building. Tasmania DoE also invited educators, support staff, and system leaders to develop the model and identify limitations. At a later stage in the project, when the project Tasmania DoE will develop a training program to show model users how to use the system predications appropriately, how to give feedback to improve the system, and to clearly outline the model’s limitations.

Human in-the-loop: At no time should educators or support staff rely solely on the output of the AI system to make decisions about student well-being, broadly construed. The AI system will always be limited by both the data sources it incorporates and by the changing conditions of well-being in local school, family, and student contexts. The system will not have all the data for perfect predictions. Only school support staff and Tasmania system leaders will suggest actions based on system outputs.

**What means will be built into the system for correction and model feedback by those who provide data and who use its outputs?**

Dashboard design – Tasmania will restrict the data surfaced by the dashboard, including, but not limited to, sensitive information that Tasmania educators and stakeholders in the system should not see or otherwise have access to. The dashboard should present the analysis the educators and support staff need while limiting their control over using the AI system in unintended ways.

Feedback options should be part of the dashboard experience for end-users to provide feedback to the system owners as recommended by the Model Recalibration section of the System Design. In addition, these feedback options will be part of the training to end-users in schools.

Privacy and Security

Private or personal data should not be collected or incorporated in analytics or AI products for education unless all groups have agreed this data is necessary to achieve the shared purpose of a specific analytics or AI project. Additionally, the people providing the data need to give permission for the data to be used for this purpose, such as through school policy at enrollment. Data providers should directly understand the value that they will receive as a result of sharing their data. Finally, the security of that data must be protected, guidelines or policies developed for which roles can access which data, and the level of anonymization needed for specific use case purposes defined.

**Developing Classifications for Datasets.**

Identifying sensitive data, such as personal information, should be part of the use case process. In OEA modules for individual datasets, sensitive data is often pre-identified, and scripts are written to pseudonymize or anonymize specific data fields before they “land” in Stage 2 data lakes and are accessed by researchers or data scientists. For datasets that are not OEA modules, the process of identifying data for sensitivity classification should be conducted through a collaboration between the project’s data engineers and individuals who understand the local education context and datasets.

**How will access to sensitive data be secured and protected in the data environment?**

*For example, is role-based access control defined and operationalized through* [*Azure Active Directory*](https://docs.microsoft.com/en-us/azure/active-directory/roles/custom-overview)*?*

Role Based Access is defined and operationalized through Azure Active Directory. Azure Purview was additionally used to identify sources of potentially sensitive data before it reached areas used for modeling and discovery in an eyes-off fashion. The use of sensitive sources of data can additionally be monitored by Purview to ensure that any transfer or use of sensitive data in other environments is monitored and subject to the same role-based access controls.

**Does the dataset contain any end-user identifiable information (EUII) and how will that data be protected and governed?**

The data contains some personally identifiable information. In no case did data scientists have access to the data outside of the role-based environment (managed by the Department of Tasmania). This data is governed using the Purview management tool for identifying and tracking all EUII Data. All EUII data was pseudonymized before it was provided to any of the data scientist tier of RBAC.

Accountability

Accountability requires that people who develop and deploy AI systems be held responsible for how they operate. AI systems should never be left to operate unchecked, irrespective of the degree to which they may be capable of acting autonomously. This is what is meant by the phrase “humans in the loop.”

**How will the analytics or AI system be monitored over time to ensure analytics and prediction perform reliably? Who will be responsible for this?**

Implementation of this predictive model should include humans in the loop in all cases as recommended by the System Design. As such, when an output prediction (e.g., a student is identified as potentially needing support and should be nominated for human assessments) made by the AI System, this prediction should first be reviewed by Student Support teams in each school. They should assess that student’s data and decide if action is needed. If the support team decides the prediction is incorrect, they need to have a means to submit this decision back to the model system. Using this information, errors in the model can be tracked, updated, and used to improve the system. This type of model feedback data is essential to maintaining and improving the model quality over time.

Inclusion

The datasets used in learning analytics and AI determine the insights and predictions produced. If those datasets do not represent the whole population of learners, if the data quality is poor, or if certain types of data are not included in the models, it will decrease the accuracy, validity, and inclusiveness of the insights. Similarly, if the way in which the insights are acted upon by the system does not include all groups (e.g., students with disabilities), it can reinforce exclusion from learning opportunities.

**How does data collection ensure that data inputs are provided by all relevant populations, including diverse or traditionally marginalized groups?**

As a result of this modelling work in Tasmania, the Department of Education mandated that all students should be assessed for well-being by staff trained in the assessment’s use to improve data quality. We recommend that other users of the package ensure all students are included as well in the data collection process.

**How will the analytics or AI outputs from the system be provided to all relevant populations, including diverse or traditionally marginalized groups?**

The results of this model will be applied equally to all students in the system. In order to ensure that results are served equally, the Department of Education is ensuring equal training on the system for education stakeholders who would are responsible for applying this tool throughout the school system. We recommend that other school systems should make their best efforts to ensure that their model systems be as inclusive as possible and employ fairness assessment for sensitive demographic groups.

**More on the OEA Use Case Template and Documentation:**

The OEA Use Case Template can be used to plan a single use case or multiple times to develop an inventory of use cases, such as when an education system is developing a comprehensive plan for data modernization.

Sections 1-3 of the OEA Use Case Template and the section on Privacy and Security in Section 4 can be used to develop any type of data use case from simple reports to more complex AI models. Completing these sections can help prevent many common problems in data projects such as:

* Asking the wrong questions or not fully understanding the problem to be solved with data
* Using the wrong type of data or too much data to solve the problem of the use case
* Making incorrect assumptions about the data and how it maps to the problem
* Developing a data solution that is not utilized by key groups for its intended purpose (e.g., not used to make decisions by schools, educators, students, families).

Section 4 should be used throughout the use case development process to operationalize and document decisions made for each of the principles of Responsible AI. This section is especially important when a use case involves the development of a machine learning model or a predictive algorithm, as these have the potential to cause unintentional harm to students.

The Use Case Template should generally be managed by the Project Manager for any specific use case, with input and review by all roles and key groups involved in use case planning.

*This is a preliminary document and may be changed substantially prior to final commercial release of the software described herein.*

*The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication.*

*This template is for informational purposes only. Microsoft makes no warranties, express or implied, in this document.*

*Complying with all applicable copyright laws is the responsibility of the user.*

*© 2021 Microsoft Corporation. All rights reserved.*