OEA Chronic Absenteeism

Use Case Defined

March 2022

The OEA Use Case Template from Microsoft Education is intended to help education systems and institutions define and plan specific data and AI projects.

This Use Case for a Predictive Model of Chronic Absenteeism was developing through a partnership between Microsoft Education, Kwantum Analytics, and Fresno Unified School District in Fresno, California.

1) The Use Case Problem

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

Student absenteeism is a fundamental challenge to education systems, and it permeates education policy and practice discussions even more as result of the global pandemic. It is a crucial time to take stock of what we know and explore new ways of addressing this problem.

Machine learning offers the potential to find patterns of absenteeism across student demographic, attendance, engagement, achievement, and social-emotional measures to predict students at risk of becoming chronically absent so that proactive and preventative interventions can be utilized to support students.

Education systems are responsible for addressing student absenteeism and can use a predictive model to focus resources to better support students on a trajectory leading to chronic absenteeism, identify the best interventions to prevent absenteeism, and ultimately reduce absenteeism.

There is a growing body of research, described below, substantiating what most parents and teachers have long believed to be true: School truancy undermines the growth and development of students. Students with more school absences have lower test scores and grades, a greater chance of dropping out of school, and higher odds of future unemployment. Absent students also exhibit greater behavioral issues, including social disengagement and alienation. The most recent national estimates in the US suggest that approximately 5–7.5 million students, out of a K–12 population of approximately 50 million, are missing at least 1 cumulative month of school days in a given academic year, translating into an aggregate 150–225 million days of instruction lost annually.

2) The Use Case Stakeholders

Who are the stakeholder groups for this use case, and how are they involved in its development?

Education system teams responsible for addressing chronic absenteeism can collaborate with technology and data groups in the system and external education analytics companies to develop a predictive model that reliably predicts students at risk of becoming chronically absent. To build an informed, ethical and effective use case, many stakeholder groups should be involved in the design and development of the use case.

Stakeholder Groups	Relationship to Use Case	Involvement in Use Case
Students	***Indirect:*** *providers of data to the predictive model, and ultimately receive preventative solutions if they fall in high-risk categories.* *They or their families or guardians should have awareness and give permission for their data to be used.*	*In the initial phases of model development and intervention designs, various types of students should be consulted in the model design process (review of data sources used, theory development). At a later stage, students identified as at-risk will receive or participate in intervention solutions and provide feedback to the system.*
Parents or Guardians	***Indirect:*** *May be providers of data to the predictive model, and ultimately receive preventative solutions if they fall in high-risk categories.* *Families or guardians should have awareness and give permission for students' data to be used.*	*In the initial phases of model development and intervention designs, parents or guardians should be involved in the model design process (review of data sources used, theory development). At a later stage, families identified as having at-risk students may receive or participate in intervention solutions and provide feedback to the system.*
Educators (Faculty or Teachers) and School Support Staff	***Indirect:*** *May participate in developing and implementing preventative solutions.*	*Educators may directly provide data or feedback to the model and utilize data or insights if they are part of an intervention to prevent absenteeism.*
School or Department Leaders	***Indirect:*** *participate in developing and implementing preventative solutions.*	*Leaders would directly utilize a data tool that reliably predicts students at risk of becoming chronically absent.*
School System or Institutional Leaders	***Direct:*** *responsible for addressing chronically absent students in schools.*	*Will lead efforts to develop the model and implement preventative solutions.*
Researchers	***Direct:*** *research Chronic Absenteeism patterns in the system and be key partner in developing the model.*	*Responsible for maintaining and updating the system to ensure ongoing accuracy of the predictive model.*
Potential Malicious Actors	***Indirect:*** *Student hackers, external hackers.*	*Corrupt data sources or modify predictive model so the model does not accurately predict at risk students. Act to misuse intervention solutions.*

Outline how stakeholders will be involved in the development in different stages of the use case development:

Early Stages: Defining the use case problem, developing the local theory or conceptual model of the problem, identifying key data sources to include in the use case in the local context:

Students, families or guardians, educators, school leaders, system leaders and researchers: In the initial phases of model development and intervention designs, these stakeholders will be involved in the model design process by providing their perspectives on causes of absenteeism in the local context (theory development), and in reviewing the data sources that are intended to be used in model development, for example to provide input on the quality and applicability of those data sources to the use case.

Focus group discussions should take place with these groups to assess their interest, concerns and ideas about this model development and the potential intervention solutions that might be valuable for the education system to provide to prevent Chronic Absenteeism.

Reviewing and Designing Outputs Stages: Testing validity of the use case results, developing dashboard designs or set of interventions based on the use case results:

As the predictive model is developed, these same stakeholders will again review the model and check for transparency, accountability, and to address other questions or concerns around how the model addresses responsible AI principles (see below). In addition, they will be asked for input on 1) how the model outputs should be communicated (e.g. dashboard designs) and; 2) the set of interventions developed to prevent Chronic Absenteeism and whether some of these interventions can or should be automated. Finally, when the model starts to be used, they should provide continuous feedback on the system, correcting the model over time.

What type of outputs are expected from this use case, such as AI models, dashboards, or notification systems?

Stakeholder Group	Outputs
Students	*Depending on the output results, students at risk may receive interventions such as support groups, medical or mental health supports, transportation assistance, dependent on the reason for their absence patterns.*
Parents or Guardians	*Depending on the output results, families of students at risk may receive interventions such as support groups, medical or mental health supports, transportation assistance, dependent on the reason for their absence patterns*
Educators (Faculty or Teachers) and School Support Staff	*Depending on the output results, educators may have access to a tool, dashboard, or data set that identifies at risk students in their current classes, the reasons that may be causing the risk, and recommends a specific intervention or provides intervention suggestions to the educator to choose among.*
School or Institution Leaders	*Depending on the output results, leaders may have access to a tool, dashboard, or data set that identifies at risk students in their schools and recommends a specific intervention or provides intervention suggestions for each student.*
System Leaders	*Data analysis and exploration dashboards to understand patterns of chronic absenteeism, changes in causes of absenteeism over time, and analysis of the impact of interventions that reduce or prevent absenteeism.*

3) Mapping Theory to Data

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

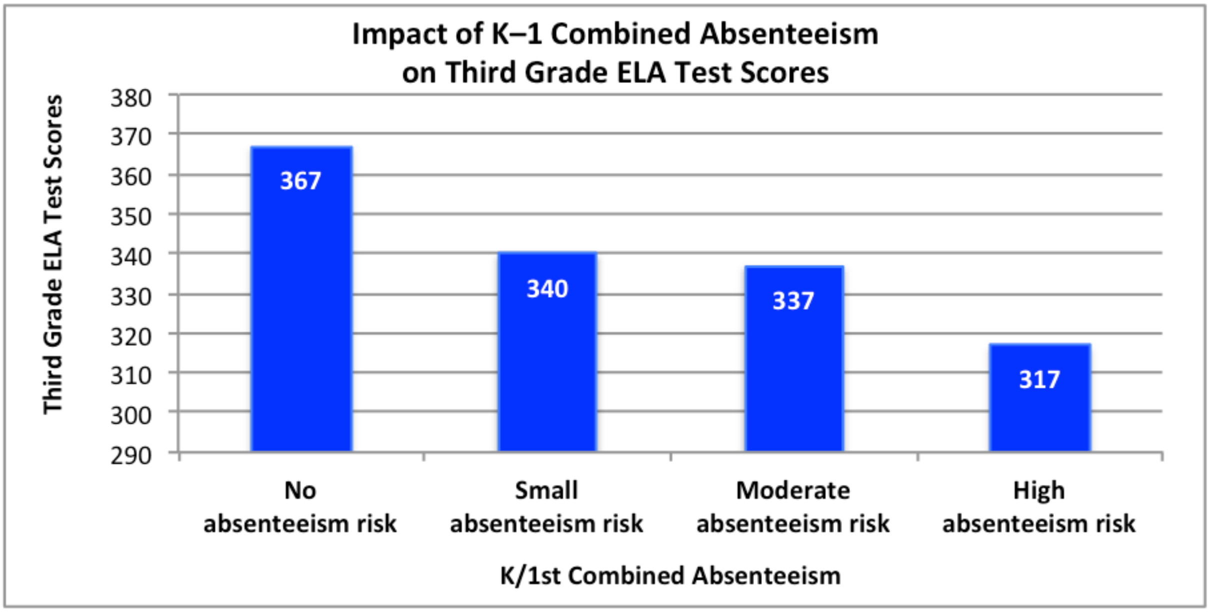
Decades of research document the significant negative impacts of student absenteeism on academic achievement, emotional development, graduation, health, and long-term success (Gottfried, 2015). Yet, until just a few years ago, the U.S. K–12 education system was virtually unaware that it had a chronic student absenteeism problem. Prior to that time, chronic absenteeism was never tracked by school systems, let alone addressed. A recent analysis of the data revealed that a significant number of students (one in seven) were chronically absent, defined as missing 10% of school days (Balfanz & Brynes, 2012).

It sounds like circular reasoning, but one of the main impacts of chronic student absenteeism is that it leads to more chronic student absenteeism. A history of chronic absenteeism is a significant predictor of future absenteeism (London, Sanchez, & Castrechini, 2016). For example, a student who is chronically absent in kindergarten is 24% more likely to be chronically absent in first grade. A student chronically absent in kindergarten, first grade, and second grade is 41% more likely to be chronically absent in third grade (Bauer et al., 2018).

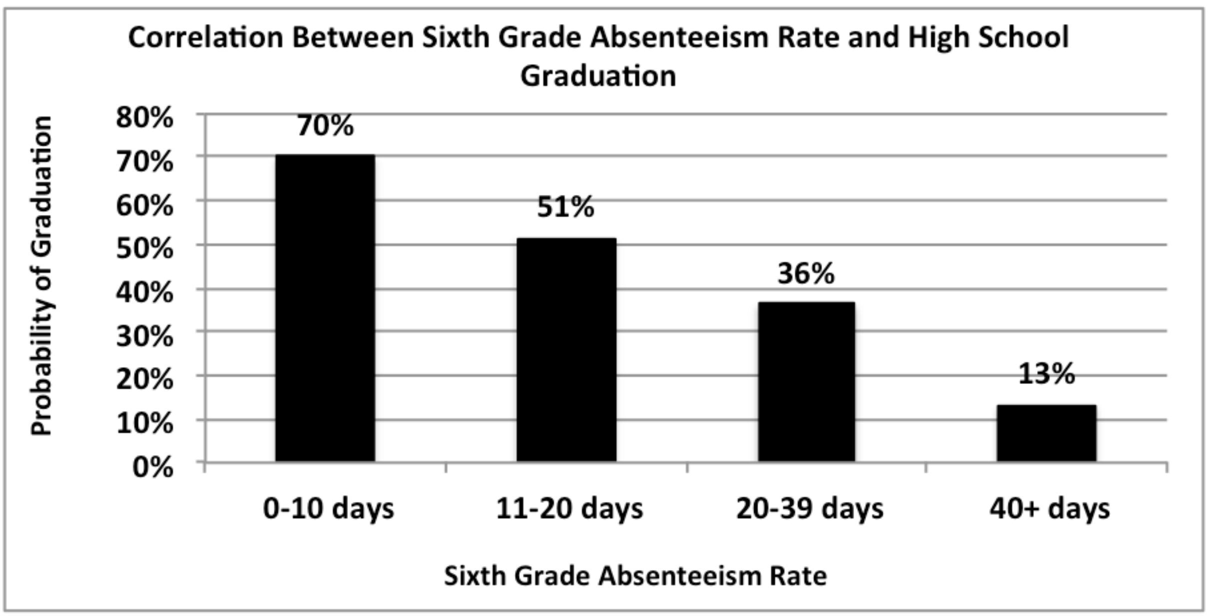
The relationship between chronic absenteeism and academic performance is evident at all grade levels, across subjects, and across assessment tools. And the relationship is always linear. Every increase in absenteeism correlates with lower academic performance.

The negative correlation between high student absenteeism and poor academic performance begins in kindergarten. There is a common belief that missing school at this age doesn't matter (Robinson, Lee, Dearing, & Rogers, 2018). Often, kindergarten is not even part of state compulsory attendance laws (which typically do not start until children are older) or is offered only half-days in many locales. Yet, there is strong correlational evidence suggesting that high absenteeism rates in kindergarten predict negative academic performance in later grades and lower high school graduation rates. These early elementary school years are critical for developing social and academic skills as building blocks for future learning. Disruptions or delays in learning these skills can have a ripple effect on all future learning (Coelho, Fischer, McKnight, Matteson, & Schwartz, 2015).

A critical benchmark in early elementary education is reading proficiency in third grade. It is an important pivot point where students shift from learning to read to reading to learn. Interventions for struggling readers after third grade are seldom as effective as those in earlier years (Fiester, 2010).



Chronic absenteeism is one of the top early indicators of students at risk of not graduating high school (Baltimore Education Research Consortium, 2011). Even absenteeism in early grades can predict at-risk students.



Chronic student absenteeism does not occur in a vacuum, and it can have a negative impact on the academic performance of all students, not just those who are absent. The educational experiences of children who attend school regularly can be diminished when teachers divert their attention from the class as a whole to meet the learning and social needs of children who miss substantial amounts of school (Chang & Romero, 2008).

The American Academy of Pediatrics recently issued a policy statement linking school attendance and good health (Allison & Attisha, 2019). After an extensive review of the evidence, the academy identified both short- and long-term health risks associated with chronic absenteeism. In the short term, the act of missing school is linked to increased unhealthy behaviors, including alcohol consumption, drug use, smoking, and risky sexual behavior. Teenage pregnancy, violence, unintentional injury, and suicide attempts are also associated with chronic absenteeism.

When it comes to chronic absenteeism, children from low-income families face triple jeopardy:

- They often live in conditions that contribute to high rates of chronic absenteeism—lack of access to adequate health care, decent housing, food, clothing, family support, and transportation (Ready, 2010).
- They are much more likely to suffer from multiplier conditions—for example, changing schools midyear, homelessness, and suspensions—that exacerbate chronic absenteeism. Each condition also has its own negative impact on student achievement (Ready, 2010).
- The poverty-related conditions make it more difficult for these students to recover from lost school days as they lack resources to help them make up for the missing time. They suffer the highest rate of loss per individual absence (Ready, 2010).

A multi-tiered tiered system is a proven model for addressing system-wide chronic absenteeism issues with students of various needs and capabilities (Kearney, 2016; Kearney & Graczyk, 2014). The model is based on the assumption that there are different categories of need requiring different levels and intensities of intervention. It is cost-effective, makes good use of limited resources, and is customized to address specific school and student needs. There are numerous versions of this type of model, but the more developed and tested models include universal prevention, early intervention, and specialized supports. (Attendance Works, 2018).

The following is a more detailed list of the types of contributors to absenteeism (Attendance Works and Everyone Graduates Center, 2018). It is not meant to be all-inclusive.

Barriers

- Illness, both chronic and acute
- Lack of physical health, mental health, vision, or dental care
- Trauma
- Unsafe path to and from school
- Poor transportation
- Frequent moves or school changes
- Involvement with child welfare or juvenile justice system

Negative School Experiences

- Struggling academically or socially
- Bullying
- Suspensions and expulsions
- Negative attitudes of parents due to their own school experience
- Undiagnosed disability
- Lack of appropriate accommodations for disability
- School climate
- Facilities (condition of the school building)

Lack of Engagement

- Lack of culturally engaging instruction
- No meaningful relationship with adults in school
- Stronger ties with peers out of school than in school
- Failure to earn credits/no future plans
- Many teacher absences or long-term substitutes

- Ineffective teaching

Misconceptions

- Absences are only a problem if they are unexcused
- Missing 2 days a month doesn't affect learning
- Sporadic absences aren't a problem
- Attendance only matters in the older grades
- Kindergarten is optional

Allison, M. A., Attisha, E., & AAP Council on School Health. (2019). The link between school attendance and good health. Pediatrics, 143(2). e20183648

Attendance Works and Everyone Graduates Center. (2017). Portraits of change: Aligning school and community resources to reduce chronic absence. Retrieved from https://www.attendanceworks.org/portraits-of-change/

Balfanz, R., & Byrnes, V. (2012). The importance of being in school: A report on absenteeism in the nation's public schools. Baltimore, MD: Johns Hopkins University Center for Social Organization of Schools.

Baltimore Education Research Consortium. (2011). Destination graduation: Sixth grade early warning indicators for Baltimore city schools. Their prevalence and impact. Baltimore, MD: Author.

Bauer, L., Liu, P., Whitmore Schanzenbach, D., & Shambaugh, J. (2018). Reducing chronic absenteeism under the every student succeeds act. The Hamilton Project. Washington, DC: Brookings Institute. Retrieved from http://www.hamiltonproject.org/assets/files/reducing_chronic_absenteeism_under_the_every_student_succeeds_act.pdf

Chang, H. N., & Romero, M. (2008). Present, engaged, and accounted for: The critical importance of addressing chronic absence in the early grades. New York, NY: National Center for Children in Poverty.

Coelho, R., Fischer, S., McKnight, F., Matteson, S., & Schwartz, T. (2015). The effects of early chronic absenteeism on third-grade academic achievement measures. Madison, WI: Robert M. La Follette School of Public Affairs, University of Wisconsin.

Fiester, L. (2010). Early warning! Why reading by the end of third grade matters. Kids Count special report. Baltimore, MD: Annie E. Casey Foundation.

Gottfried, M. A. (2015). Chronic absenteeism in the classroom context: Effects on achievement. Urban Education, 54(1), 3–34.

Kearney, C. A. (2016). Managing school absenteeism at multiple tiers: An evidence-based and practical guide for professionals. New York City, NY: Oxford University Press.

Kearney, C. A., & Graczyk, P. (2014). A response to intervention model to promote school attendance and decrease school absenteeism. Child & Youth Care Forum, 43(1), 1–25.

London, R, A., Sanchez, M., & Castrechini, S. (2016). The dynamics of chronic absence and student achievement. Education Policy Analysis Archives, 24(112), 1–27.

Ready, D. D. (2010). Socioeconomic disadvantage, school attendance, and early cognitive development: The differential effects of school exposure. Sociology of Education 83(4): 271286.
Attendance Works. (2018). 3 tiers of intervention. Retrieved from https://www.attendanceworks.org/chronic-absence/addressing-chronic-absence/3-tiers-of-intervention/

Robinson, C. D., Lee, M. G., Dearing, E., & Rogers, T. (2018). Reducing student absenteeism in the early grades by targeting parental beliefs. American Educational Research Journal, 55(6), 1163–1192.

Mapping Theory to Data. From prior research or conceptual models what are they key data categories expected to inform this use case? What local data sources are available or needed for each category? Please note where no data is available for a Data Category

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. Identifying which data should be viewed as a "feature" and which data is the "target outcome" is at the core of this mapping.

Key Data Category	**Local Data Source**
1. *For example, Attendance Data*	*For example, Student Information System – "Days Absent"*
2. *For example, Digital Learning Application Data*	*For example, Canvas and M365 app usage data*
3. *For example, Student Well-Being*	*For example, Student Health Records*
4. *For example, Student Behavior*	*For example, Student Behavior records from the Student Information System*

Note: OEA modules can provide data sources that support the chronic absenteeism use case through accelerating the ingestion of key data sources needed and providing resources to set up these use cases.

Note: Mapping theory to data with a 'data dictionary.'

A "data dictionary" allows the data team to examine specific data tables and data entities in the available datasets, and then map specific items to the Key Data Category.

New data services like Azure Purview can support this work through creating a holistic, up-to-date map of a data repository with automated data discovery, sensitive data classification, and end-to-end data lineage.

Please see "Privacy and Security" section below for more ensuring that sensitive data is protected.

<table>
<tr><th>Theoretical Construct Data</th><th>Local Data Source Mapped to Theoretical Construct</th></tr>
<tr><td>1. Barriers
Illness, both chronic and acute
Lack of physical health, mental health, vision, or dental care
Trauma
Unsafe path to and from school
Poor transportation
Frequent moves or school changes
Involvement with child welfare or juvenile justice system</td><td>This is where specific fields from available datasets should be mapped to the theoretical constructs important to Chronic Absenteeism.</td></tr>
<tr><td>2. Negative School Experiences
Struggling academically or socially
Bullying
Suspensions and expulsions
Negative attitudes of parents due to their own school experience
Undiagnosed disability
Lack of appropriate accommodations for disability
School climate
Facilities (condition of the school building)</td><td>This is where specific fields from available datasets should be mapped to the theoretical constructs important to Chronic Absenteeism.</td></tr>
<tr><td>3. Lack of Engagement
Lack of culturally engaging instruction
No meaningful relationship with adults in school
Stronger ties with peers out of school than in school
Failure to earn credits/no future plans
Many teacher absences or long-term substitutes
Ineffective teaching</td><td>This is where specific fields from available datasets should be mapped to the theoretical constructs important to Chronic Absenteeism.</td></tr>
</table>

4) Responsible AI Principles Applied

In these next sections, please answer the questions under each of the headings describing how responsible AI principles will be applied to this use case.

Fairness Principle

Who is most likely to be at risk of experiencing harms from this use case?

"Harm" can be subjective. Children from low-income families and/or children in traditionally underserved student groups (student with disabilities, English language learners, youth involved in the juvenile justice system) typically have higher incidences of absenteeism and therefore are deserving of additional support. On the other hand, parents and community leaders representing low-income and/or traditionally underserved student populations are extremely sensitive to systems that profile and/or track students and are not easily convinced that such systems are in the best interests of their children.

Parents and community leaders can also be highly skeptical of systems that are not easily explainable.

- Removing a teacher from a specific post (based on false or misleading information)
- Supports for a student withheld (in false negative)
- Misdiagnosing causes of student absenteeism – and then providing them with the wrong or inappropriate supports

Planned Mitigations:

- Involve stakeholders from each of the groups in the early planning, design, development and testing of the model and interventions
- Pilot slowly and iteratively test to improve accuracy of predictive model
- Build feedback loop into system for student, teacher, and all stakeholders in the system

Group or Subpopulation	Clearly Labelled in Dataset? Y/N	Planned Mitigations
1. *For example, Immigrants*	*Y*	*For example, Test model outputs via Fairlearn for false positives / negatives*
2. *For example, Rural Students*	*N*	*For example, Add geolabel based on address mapping for rural, then test outputs via Fairlearn*

3. *For example, Low Income Students*	*Y*	*For example, Test model outputs via Fairlearn for false positives / negatives*
4. *Students who are close to becoming chronically absent or have just become a different level of absenteeism*	*Y*	*For example, risk of absenteeism is fluid and changes frequently.*

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?

The predictive model for Chronic Absenteeism is least accurate for "edge" cases. Because it is challenging to differentiate a student with 91% attendance from a student with 89% attendance, accuracy of such predictions will be lower. In contrast, the model will be better at predicting for students who have very high attendance (near 100%) and very low attendance (say 70% attendance or less).

A second risk factor is model disparity in relation to student demographics. Even though demographic data such as gender or race are not included in the model building process, the model may interpret demographic separations indirectly through other variables such as income level or exam scores. Because of this, the model may perform better for some demographic groups than others. This principle is known as "model fairness".

Planned Mitigations:

To protect against a model performing better for some groups of students compared to others, 2 key strategies need to be incorporated.

1. *Monitoring* of model performance across different groups of students is needed. This can be done with a basic PowerBI dashboard. It is essential that any model performs fairly for all demographic groups. For target prediction of chronic absence, it may be useful for stakeholders to want to model to perform most accurately on actionable student groups such as students who are most at-risk.
2. *Retraining* models which perform fairly. Two approaches here include balancing training data across demographic groups of concern (ie Fairlearn) and building complex enough models (such as ensembles) which are rigorously tested via cross validation practices.

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.

What steps will the analytics or AI process include?

The AI process is outlined in 4 main steps:

1. *Data subsetting and aggregation:* Data identified in the above theory and data discussion will be located in the production data environment. Only the columns needed for model building will be used.
2. *Feature engineering and model table construction:* Data from step 1 will be combined into a single table for building the predictive model. Because each row in this table represents 1 student's data, certain columns such as attendance records will need to be aggregated into a single metric.
3. *Model building:* AutoML will be used to build a best performing model via Azure Machine Learning Studio. This process will record and catalog the model table and all models produced for future reference. The model will be used to make predictions on the model table and InterpretML will be used to identify individual feature improtance.
4. *PowerBI deployment*: Model results and other data (ie SIS, time dependent attendance, demographics) will be made queryable by a PowerBI dashboard. This dashboard will be used to explore model findings and assess fairness.

Who will develop the analytics or models?

In the Fresno Unified Predictive modelling work, Kwantum Analytics developed the primary predictive model, working closely with Fresno Unified School District leaders and Microsoft Education.

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

Training should be conducted for all users and stakeholders of the Chronic Absenteeism model, including training on the limitations and weaknesses of the model. In particular, training should include guidance on how they can use the system to inform interventions or actions taken with absent students, but that the system should be used to inform their decisions, not make their decisions. If the end user's judgement of a student's situation does not match the system's data, the end user should be trained to provide feedback to better train the model, or at least question the data being presented.

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

As part of the above training, users of the model:

1) Should provide feedback or additional inputs to the system (such as tagging which interventions or activities were taken)
2) Should provide feedback to the system if the data or recommendations were incorrect or questionable given the user's knowledge of the student and their context.

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. Ideally, 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.

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.

1) Risk 1: Data about levels of student absenteeism and the causes of that absence (e.g. medical) become publicly available or available to individuals others than those who should have role-based access to such information.

2) Risk 2: Data about the health and wellbeing of students, families, educators, or staff is unsecured and becomes associated with the students' personal information.

3) Risk 3: Data about interventions taken with students and families becomes available to individuals or the public who should not have access to such information.

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

Only pseudonymized student data will be used to build and assess models, so no students will be identified through this process. If FUSD wishes to identify students, such work will be performed internally with appropriate permissions governed through Azure Active Directory.

Does the dataset contain any personally identifiable information (PII) and how will that data be protected and governed?

Only pseudonymized student data will be used to build and assess models, so no students will be identified through this process. If FUSD wishes to identify students, such work will be performed internally with appropriate permissions governed through Azure Active Directory.

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.” A part of this is ensuring documentation of the decisions made during the AI system development. This document can be used for that purpose.

Who is responsible for reviewing the Use Case documentation and ensuring that the implementation meets responsible AI principles?

The decision makers which will use this research’s findings in practice will be responsible for knowledge of responsible AI principles and monitoring of AI predictions according to what was outlined above.

How will stakeholders and end users be trained on the appropriate use of the system?

Kwantum Analytics will train key stakeholders on interpretations of model accuracies via deployed PowerBI dashboards. Dedication documentation will be created as a refence.

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

Kwantum Analytics will be responsible for monitoring analytics and prediction performance during the initial stages of this work. Through communications with stakeholders, an accuracy threshold (ie 80% accurate) will be determined to decide if model quality is sufficient for actionability. Model accuracy should be checked quarterly with the use of new attendance records. It is recommended that the model be retrained every 1-2 years by either FUSD data scientists or and outside partner.

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 the insights are acted upon by the system do not include all groups (e.g., students with disabilities), it can reinforce exclusion from learning opportunities.

What are the constraints of these local data sources for this specific use case?

Please describe the limitations of these datasets. For example, are the datasets missing data for certain student populations? Is there bias in the data collection method? This will inform the sections below on Responsible AI principles.

Dataset Name	Constraints or Limited Representativeness

1. Student absence codes	*For many students, the absence code reported by schools is "No Clearance" without detail on why the student was absent. It is unknown whether such students are absent due to medical, transportation, or other reasons, which might improve the predictive accuracy of the model and provide better representation.*
2. For example, digital learning app data	*For example, how does data collection ensure that data inputs are provided by all relevant populations, including diverse or traditionally marginalized groups? Do all students have digital access?*
3.	
4.	