CSE 163 Term Project

Title and author(s): Predicting Kindergarten Readiness of children based on different socio-economic identities of children.

<u>Author(s)</u>

David Zoro, Gita Dhungana, and Aynur Sari

Summary of Research Questions:

1) In what domains are students most likely to be prepared for?

2) What group of students are more prepared for X domain (among the selected domains) for kindergarten? We are selecting following domains:

   a) Cognitive

   b) Social - Emotional

   c) Literacy

3) Are there any correlations between kindergarten preparedness in the above three domains and the following student groups:

   a) Migrant vs non-migrant (isMigrant)

   b) English language learners vs non-English language learners (Bilinguals)

   c) low income vs non-low income (FreeLunch)

Motivation:

A child's brain develops rapidly in the first few years of their life. Studies have found that infant brains overproduce neuron connections, and as children grow, the excess connections that are not activated (stimulated enough) begin disappearing in a process called *pruning*. Various environmental factors (such socio-economic backgrounds) can affect a child's physical, emotional and cognitive growth. A child's developmental readiness for kindergarten is vital for them to flourish and prosper in kindergarten and beyond.

In the United States, children are enrolled in kindergarten when they turn five years old. However, according to the Education Research and Data Center's Early Learning Feedback report published in 2018 (and updated in 2021), many children enter kindergarten without the foundational skills they need to be successful. According to Washington State Department of Children Youth and Families about 80,000 children enter kindergarten in Washington's public school system every year, and only about half of children (across all income groups) are ready for kindergarten based on the state's kindergarten readiness assessment. The assessment is called ***Washington Kindergarten Inventory of Developing Skills (WaKIDS) whole-child assessment.*** Children, who start kindergarten behind their peers, need appropriate support to close the achievement gaps as early as possible.

In this project, we examined a dataset about children's kindergarten preparedness in Washington (collected through WaKIDS assessment over a period of seven years) to see how different social and economic identities of children are related to their preparedness in six skill areas.

We created three different machine learning models based on the dataset, and tested the models to evaluate their accuracy. These models are useful to predict children's preparedness in the given skill areas based on their social identities. The three models were built based on SGDRegressor, sklearn.linear_model.LinearRegression and DecisionTreeRegressor algorithms. After evaluating these models based on their mean squared error and R-squared value, we conclude that DecisionTreeRegressor is the most optimal model for our dataset with R-squared value of 1 and mean squared error of 0.0068.

In addition to a machine learning model, we also showed the relation between a child's different social identities and their development in certain skill areas through data visualizations.

Dataset:

Report Card Kindergarten Readiness for 2011-12 to Most Recent Year

We retrieved the dataset from Data.Gov in May 2022. The data was collected through WaKIDS whole-child assessment of kindergarten-going children carried out at the beginning of every school year in Washington state. According to WaKIDS FAQ page all districts that provide state-funded, full-day kindergarten education in Washington are legally required to assess incoming kindergarten students at the beginning of each school year.

The assessment involves kindergarten teachers examining children’s developmental readiness using an observational assessment tool that measures children’s skills and abilities in six skill areas: ***social-emotional, physical, language, literacy, cognitive, and mathematics.*** In certain school districts and in some years, ***Spanish language and Spanish literacy*** are added to the list, increasing the total number of skill domains to eight.

The dataset includes data from 2013-14 to 2019-20 school year. The data is grouped based on different social identities of the children as well as by the school districts. Since it is a large dataset with over 1 million rows and 44 columns, we removed several columns (*in python)* and only included six relevant columns in the final dataset. The final dataset included the following columns: ***SchoolYear, DistrictName, StudentGroupType, StudentGroup, Domain, PreparedForK_Percent.***

Of the six columns,*"PreparedForK_Percent"* is the numerical variable, and it is the outcome variable (the *label*) of our machine learning model. The variable represents the ratio of children who are prepared in a given skill area based on the WaKIDS assessment tool.

The rest of the variables (features) are categorical variables. Of the feature variables, "StudentGroupType" represents a type of social identity based upon which students are categorized (E.g., Gender or Bilingual). On the other hand, "StudentGroup" represents a particular social group under the given StudentGroupType. So for example, if the StudentGroupType is Race/Ethnicity, StudentGroup can be one of the racial or ethnic groups identified in the dataset. "Domain" refers to six (or 8, depending upon the districts and year of assessment) skill areas in which children are assessed.

Challenge goals:

Machine learning: One of our challenge goals is to use machine learning to create a regression model to predict kindergarten preparedness of children for the given skill areas. We decided to use interpretable models instead of a black box model (neural network mode). This is because although black box models are often accurate, they are hard to interpret, which makes it difficult to understand their biases and how to remove them or fix them.

New library: We train three machine learning models on our dataset that involve using two new libraries other than the ones introduced in the class. They are SGDRegressor and sklearn.linear_model.LinearRegression¶. We compared their performance with each other as well as with a model built on the DecisionTreeRegressor algorithm.

Result Validity: We used a test of statistical significance, Z-test, to see if our smaller dataset is statistically similar to our larger dataset. If it is similar, it will allow us to graph the smaller dataset and know what to expect when we graph the larger dataset. Moreover, we will use the P-value to see if we should reject the null hypothesis: the smaller dataset is statistically significant to the larger one. If the P-value is higher than alpha (0.5), we will not reject the null. Otherwise, we will reject the null hypothesis.

Method:

Machine Learning Model:

Our machine learning has to be a regression model because our outcome variable is a numerical variable.

Since our dataset is very large with over a million rows of data points, we decided to use **Stochastic Gradient Descent's Regression model class** ***(or SGDRegressor*** **)** from ***sklearn*****.** According to the documentation of the model class library, SGDRegressor is well suited for regression problems with a large number of training samples (> 10,000).

In our SGDRegressor model, we change the 'loss' hyperparameter to 'huber' from the default parameter of 'squared_error'. This modification allows the model to focus less on getting outliers correct by switching to a linear loss (from the ordinary squared error) past a distance of epsilon. By default, the epsilon hyperparameter is set at 0.1, and we kept the default value.

We also modified the 'shuffle' hyperparameter to False (from the default value of "True"), which means the training data is not shuffled after each epoch. We found that modifying the 'shuffle' hyperparameter to False improves our model's mean-squared-error and R-squared value.

Lastly, we modified the 'penalty' hyperparameter of our model to "Elastic-net" from default L2. According to the model class' documentation, this technique is useful when there are multiple highly correlated features. Given that our dataset has multiple features that are likely to be correlated with each other (e.g., "Bilingual" category and "isMigrant" category are likely to interact with each other), we selected the "Elastic-net" hyperparameter.

Also, we standardize the data set before building a model on it, by preprocessing it with StandardScaler and make_pipeline package from sklearn.preprocessing library, as suggested in the documentation of the model library.

After building the SGDRegressor model, we decided to compare its mean squared error and R-squared score with other interpretable models, namely DecisionTreeRegressor model and sklearn.linear_model.LinearRegression¶

With DecisionTreeRegressor, we found that keeping the 'max_depth' hyperparameter to its default value (None) produced lower mean squared error and higher R-squared value than modifying 'max_depth' to 2 or other values. The default value of max_depth is None, which indicates that decision tree leaves nodes are expanded until all leaves contain less than min_samples_split samples.

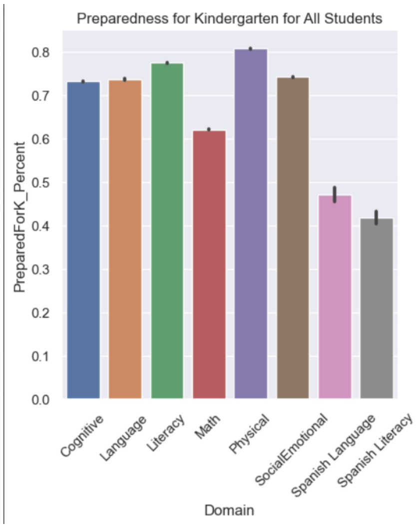
Data Visualizations and Statistics: Moreover, we created graphs to analyze the data visually and make conclusions about the preparedness of our selected groups. We filtered the data to be able to answer our research questions properly. Afterward, we dropped NaN values that would alter the visualization results. Once that was complete, we plotted the data using a bar plot. The bar plot demonstrates where each group of students are regarding preparedness. The bar plot also displays uncertainties, which is a range for our data to fluctuate on how prepared students are. We also filtered both datasets and dropped the NaN values from them. We took the average of the larger dataset as our null hypothesis. We did this to see how statistically related the larger dataset and the smaller dataset are from each other. We received a test statistic of zero and a P-value of 1, which accepts the null hypothesis.

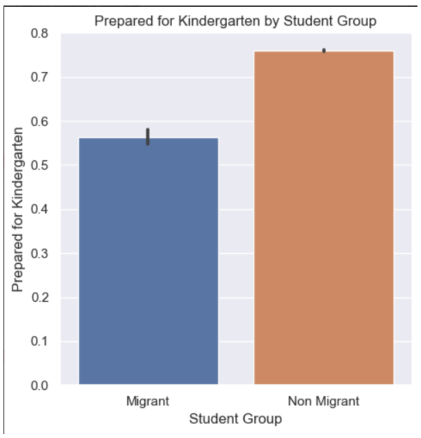
Results:

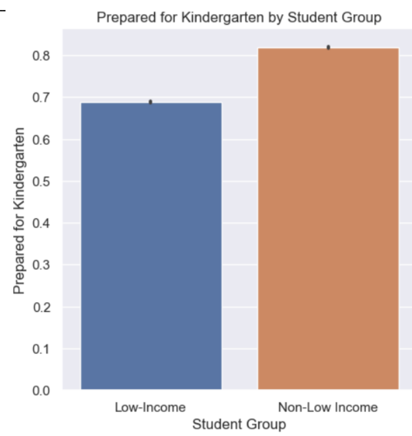
For research question one, we found that students have better preparation in the literacy and physical domains. Our result makes sense because literacy is among the first things students learn before school, and they are at the age for school. A significant finding is that students are less prepared for the math and cognitive domains. A consequence of being less ready for the cognitive or math domain is students may have to spend more time to increase competence in those areas.

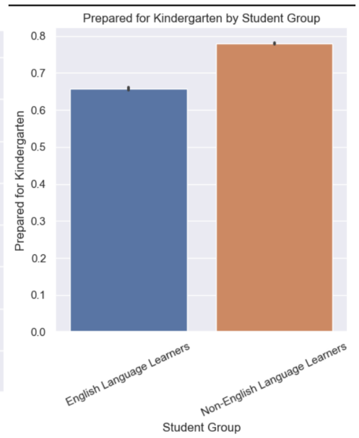
For research question two, we found that students who are non-low-income, non-migrant, and non-English language learners tend to be more prepared for kindergarten than students in disadvantaged groups. Our results are all-encompassing and take into consideration all domains. The results are important because they suggest that students from disadvantaged backgrounds are not always as prepared as students from more advantaged backgrounds. A consequence of this is that disadvantaged students may need more resources to foster equity in a kindergarten classroom. However, the results do not mean a disadvantaged student is not as capable as a student from a more advantageous background.

For research question three, we have found a correlation between kindergarten preparedness and being a member of a particular group. By using machine learning, we were able to predict a student's preparedness based on their background accurately: The relationship was an R-squared of 1 and a mean squared error of 0.0068. The analysis shows the effect a student's background can have on their academic preparedness. The accuracy of our model is important because it shows that background is correlated to student preparation. A consequence of our results is that a capable student, who is low income, could be at an initial disadvantage due to their background, which is counter to creating equity in the classroom.









Impact and Limitations

There can be various reasons why developmental gaps exist among children of different groups. And it is vital that such gaps are closed as early as possible to ensure that all children get the same access to education and other opportunities in life. Indeed, the intended goal of this dataset, according to Washington Office of SuperIntendent of Public Instruction, is to close gaps among different groups of kindergarten-going children by providing individualized support to the children in need as well as by making policy decisions that benefit all children in a particular district based on the needs of the student profile of the district.

However, machine learning models, and people alike, are likely to simplify the results of the dataset to conclude that children from certain social groups or from certain districts tend to be less capable or intelligent than the other groups. These results do not take into account the fact that socially and economically marginalized children do not have the same opportunities and resources that children from privileged families would have. For example, often the day-care cost and the cost of enrolling children in activities are prohibitively high for children coming from low-income families.

For example, the results of the dataset show that children coming from certain backgrounds (low-income families, migrant families and non-English speaking families) tend to show slower kindergarten readiness (as defined by WaKIDS assessment tools) than their counterparts from non-low income, non-migrant, and English-speaking families. Consumption of such results without critically examining the biases present in the dataset and intercultural differences among different groups could lead teachers and educators to develop implicit and subconscious biases against these groups of children and their families. This can harm these children, especially if teachers aren't aware of their biases or check their biases. It will be harmful for children if educators and policymakers lump all children of certain groups in pre-defined boxes without paying close attention to each individual child's unique situations.

These results of the dataset can also label certain school districts poorer/ disadvantaged than others. This can be helpful to a certain extent in that this helps policy makers to bring policies to improve the situations of the children there by pouring required resources to the school areas. However, on the flip side, this could turn into a vicious cycle: as certain school districts are deemed better than others based on the result of this dataset, more and more people -- who can afford -- will look to migrate to the 'better' school areas. This means there will be an imbalance in tax-based incomes between “richer” school districts and “poorer” neighborhoods, resulting in poorer neighborhoods having limited financial means to support the children living in those districts.

Interpretation of the results for children with special needs and circumstances

Designers of the WaKIDS assessment tool appear to have considered the unique situations of children with special needs and circumstances (such as children with disabilities or children from non-English speaking families etc.) by ensuring that the assessment is administered by qualified kindergarten teachers, who have taken compulsory 12-hour training on WaKIDS assessment. In addition, examiners include special needs teachers for children with disabilities and transitional teachers for children that are going through some transitions (such as recent migration). This is good news, but there still exist risks of false positives where children are deemed developmentally delayed due to their different language, and cultural and socio-economic backgrounds as well as different educational needs than the mainstream children. For example, children from families that speak a different home language than the English language may be disadvantaged when being assessed in English language. While WaKIDS appears to take this factor into account in their assessment tools, it is still important to be mindful of possibilities of such biases in the dataset.

Work Plan Evaluation:

The proposed work plan we established was good. However, we did not describe our means of communication. If we had described how we would work together, the work plan would have been better. The hours we established that we would be working together were excellent because it was a project that required member involvement. The details of what we would be doing were also excellent. We were descriptive in what we wanted to do and how.

Our proposed work plan estimates were somewhat accurate. For instance, the time we anticipated investing in the project was accurate. Our project required meetings to keep everyone up-to-date. Moreover, we need to program to answer our research questions. As our project progressed, we solidified our means of communication. It was clear to all of us how we would communicate and complete the project. The way we completed work was also somewhat accurate. However, we realized that we needed to delegate tasks to complete the project in a timely manner. If we always met at a set time to work, the project would be difficult to complete.
Our estimates were close to reality. How we worked slightly resembled what we established in our work plan. Moreover, we invested sufficient time in the project to see its completion. Our estimates accounted for the fact that the project would take time. As a brand new team, we recognized we would need to communicate and have meetings.

Testing:

Since our dataset is large, we first tested our machine learning models and data visualization programs using a smaller dataset that included data from the 2019-20 school year. We then compared the output of the smaller dataset to the output of the largest dataset.

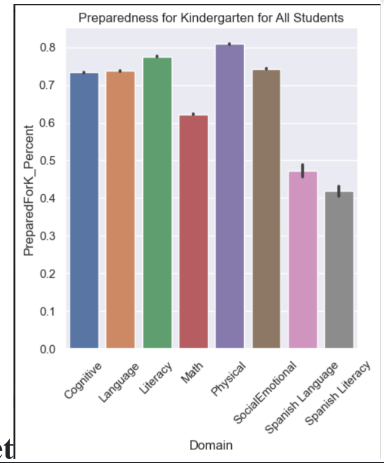
In terms of machine learning models, we found that our DecisionTreeRegressor model trained on the larger dataset performed the best out of all the three models trained on the dataset. The DecisionTreeRegressor achieved the R-squared value of 1 and mean squared error of 0.0068 when evaluating it on a test set, a subset (20%) of the total data. The model, when trained on the smaller dataset, produced on average mean squared test error of 0.009, with a R-squared value of 1

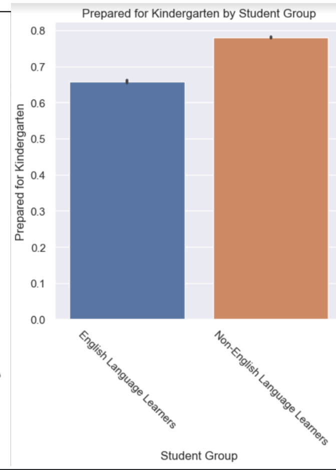
The SGDRegressor model trained on the larger data set produced a mean squared error of 0.141 and an R-squared value of 0.424. Despite the sklearn's claim that SGDRegressor has the advantage of being efficient for training on a large dataset, we found SGDRegressor was not any faster than DecisionTreeRegressor or LinearRegression models. In the smaller dataset, the SGDRegressor model produced a mean squared error of about 0.13 with an R-Squared value of 0.54.

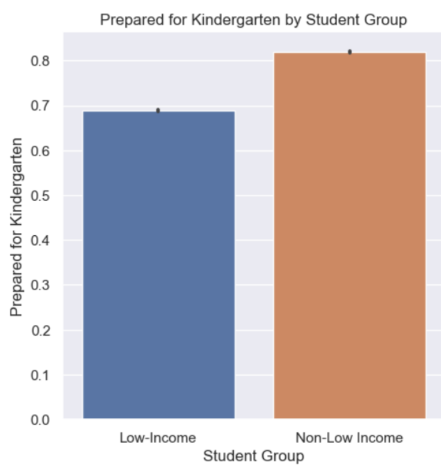
The LinearRegressor model trained on the larger dataset produced a mean squared error of 0.096 with R-squared value of 0.603. In the smaller dataset, the model produced an average error of about 0.083 with R-squared value of about 0.68.

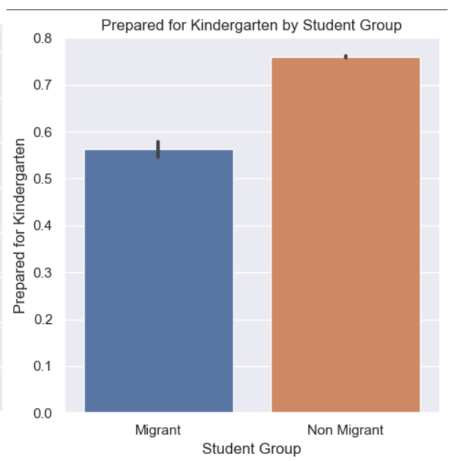
For Data visualization too, we first produced plots based on the smaller dataset, and used these graphs as expected plots to test our program in the larger dataset. Our reasoning was that regardless of the finer details, the plots should produce similar results.

Small Dataset

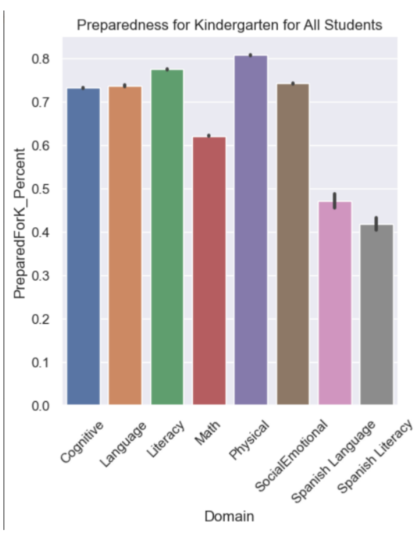


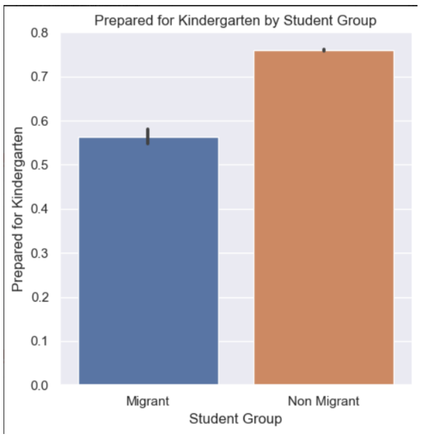


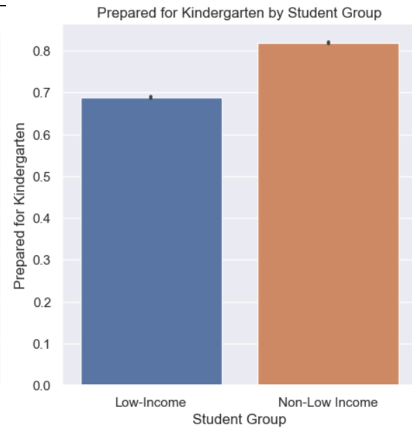


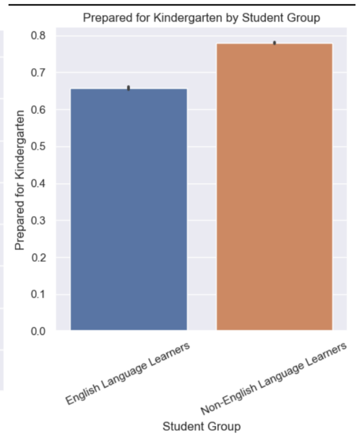


Large Dataset









Collaboration:

We referred to Washington government's publications and reports about the WaKIDS assessment and the government's utilization of the results. These are the webpages that we referred to understand the background information about the dataset and how its results are used by the government.

https://www.k12.wa.us/sites/default/files/public/wakids/pubdocs/WaKIDS_Frequently_Asked_Questions.pdf

https://www.k12.wa.us/student-success/testing/state-testing/washington-kindergarten-inventory-developing-skills-wakids

https://erdc.wa.gov/data-dashboards/early-learning-feedback-report-0

https://www.dcyf.wa.gov/node/3254